

PREFFI 2.0: Health Promotion Effect Management Instrument. Development, validity, reliability and usability

Citation for published version (APA):

Molleman, G. R. M. (2005). *PREFFI 2.0: Health Promotion Effect Management Instrument. Development, validity, reliability and usability*. [Doctoral Thesis, Maastricht University]. Netherlands Institute for Health Promotion and Disease Prevention (NIGZ). <https://doi.org/10.26481/dis.20050512gm>

Document status and date:

Published: 01/01/2005

DOI:

[10.26481/dis.20050512gm](https://doi.org/10.26481/dis.20050512gm)

Document Version:

Publisher's PDF, also known as Version of record

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

Preffi 2.0:
Health Promotion Effect
Management Instrument

Preffi 2.0: Health Promotion Effect Management Instrument

Development, validity, reliability and usability

Colofon

**Titel: Preffi 2.0: Health Promotion Effect Management Instrument -
Development, validity, reliability and usability**

Publisher: Netherlands Institute for Health Promotion and Disease Prevention
(NIGZ), Woerden

First edition: 2005

ISBN: 90-6928-240-2

Author: Gerard Molleman

Translation: Jan Klerkx

Layout: Om tekst en vorm, Utrecht

Cover design: Drukwerkstudio.nl, Harmelen

Printing: Printpartners Ipskamp, Enschede

Copyright: Oxford Pres University (chapter 3)

Copyright: NIGZ 2005/G.R.M. Molleman. No part of this publication may be reproduced, either in folio or digital form, without written permission. Requests concerning the reproduction of images or text should be addressed to:

Uitgeverij NIGZ

P.O. Box 500

3440 AM Woerden

The Netherlands

E-mail: uitgeverij@nigz.nl



Preffi 2.0: Health Promotion Effect Management Instrument

Development, validity, reliability and usability

Proefschrift

ter verkrijging van de graad van doctor
aan de Universiteit Maastricht,
op gezag van de Rector Magnificus,
Prof.mr. G.P.M.F. Mols,
volgens het besluit van het College van Decanen,
in het openbaar te verdedigen op
donderdag 12 mei 2005 om 14.00 uur

door

Gerardus Richard Maria Molleman

Promotores

Prof. dr. C.M.H. Hosman

Prof. dr. G.J. Kok

Beoordelingscommissie

Prof. dr. N.K. de Vries (voorzitter)

Prof. dr. B. van den Borne

Prof. dr. Mr. R.P.T.M. Grol

Dr. H. Schaalma

Dr. A. J.J. Voorham (GGD Rotterdam)

Het verschijnen van dit proefschrift is mede mogelijk gemaakt door het Nationaal Instituut voor Gezondheidsbevordering en Ziektepreventie (NIGZ) en subsidies van ZonMw

Original articles

The following chapters are published or submitted.

Chapter 2:

Molleman, G. R. M. and Hosman, C. M. H. (2003) Ontwikkeling van een kwaliteitsinstrument voor de effectiviteit van gvo/preventie-programma's; de Preffi 1.0, ontwikkeling en ervaringen en uitgangspunten voor een Preffi 2.0. *TSG Tijdschrift voor gezondheidswetenschappen*, 81, 238-246.

Chapter 3:

Molleman, G. R. M., Peters, L. W. H., Hosman, C. M. H. and Kok, G. J. (2004) Implementation of a quality assurance instrument (Preffi 1.0) to improve the effectiveness of health promotion in The Netherlands *Health Educ. Res.*, cyg142.

Chapter 4:

Molleman, G. R. M., Peters, L. W. H., Hosman, C. M. H., Hommels, L. M., Ploeg, M. A., Kok, H. H. and Llopis, E. J. (2003) De Preffi 2.0: systematische ontwikkeling van een kwaliteitsinstrument voor gezondheidsbevordering. *TSG Tijdschrift voor gezondheidswetenschappen*, 81, 247-255.

A modified version of this article is published in a special edition in French of Promotion & Education Hors série, October 2004 pages 22-27: Molleman, G. R. M., Ploeg, M. A., Hosman, C. M. H. and Peters, L. W. H. (2004) Preffi 2.0: un outil néerlandais pour analyser l'efficacité des interventions en promotion de santé *Promotion & Education*.

The same version is accepted for publication in the English edition of Promotion & Education, after a peer review procedure: Molleman, G. R. M., Ploeg, M. A., Hosman, C. M. H. and Peters, L. W. H. Developing a Health Promotion Effect Management Instrument: Preffi 2.0.

Chapter 6:

Submitted as: Molleman, G.R.M., Peters, L.W.H., Hosman, C.M.H. and Kok, G.J., Project quality rating by experts and practitioners: experience with Preffi as a quality assessment instrument.

Contents

Original articles	8
Contents	9
Preface	13
 Chapter 1 Introduction	 17
1.1 Background and problem definition	17
1.1.1 Focusing on effectiveness	17
1.1.2 Improving effectiveness	18
1.1.3 International experiences and approaches	22
1.1.4 Conclusion	24
1.2 Role and status of the effectiveness programme in quality management	24
1.3 Choices and requirements for a new instrument	28
1.3.1 The concept of effectiveness and the use of effect predictors	28
1.3.2 Preffi as a quality management instrument	30
1.3.3 Requirements for Preffi based on knowledge about guidelines and their implementation	31
1.3.4 Requirements for Preffi in terms of validity and reliability	35
1.4 Developing and implementing Preffi (1994 – 2004)	35
1.4.1 Principles	35
1.4.2 Developing Preffi 1.0 (1994-1995)	36
1.4.3 Implementing Preffi 1.0 (1997-2000)	37
1.4.4 Developing Preffi 2.0 (November 2000 – January 2003)	40
1.4.5 Implementing Preffi 2.0 (2003 – present)	42
1.4.6 International activities	42
1.5 General research questions and structure of this thesis	43
 Chapter 2 Preffi 1.0: development, experiences and starting points for a Preffi 2.0	 45
2.1 Introduction	45
2.2 Preffi	47
2.2.1 Background	47
2.2.2 Development	47
2.2.3 Emphasising the practitioners' perspective	48
2.2.4 Main focal points in Preffi	49
2.2.5 Preffi from an international perspective	49
2.3 A study of experiences with the use of Preffi to improve effectiveness	50
2.3.1 Questions addressed by the Preffi implementation programme	50
2.3.2 Research methods	50

2.3.3 Interventions to promote the use of Preffi	51
2.3.4 Results	51
2.3.5 Advantages of using Preffi	53
2.3.6 Disadvantages of using Preffi	54
2.4 Tension between preferred and actual situations	55
2.5 Consequences for Preffi 2.0	56
2.6 Sustained implementation and development programme for Preffi	57

Chapter 3 Implementation of a quality assurance instrument (Preffi 1.0) to improve the effectiveness of health promotion in the Netherlands

3.1 Introduction	59
3.2 Theoretical basis for the implementation process of Preffi	60
3.3 Research method	62
3.3.1 Samples	62
3.3.2 Instruments	63
3.3.3 Data analysis	65
3.4 Results	65
3.4.1 Familiarity with and use of Preffi	65
3.4.2 Attitude, support, self-efficacy and intention	68
3.4.3 Predictors of intention and use	70
3.5 Discussion	72
3.6 Practical implications	75

Chapter 4 Preffi 2.0: systematic development of a quality assurance instrument for health promotion

4.1 Introduction	77
4.2 Developing Preffi 2.0	78
4.3 Preffi 2.0	81
4.3.1 The Preffi 2.0 model	82
4.3.2 The list of criteria and changes relative to the previous version	82
4.3.3 The scoring method	85
4.4 Results of the pilot study among practitioners	86
4.5 Instructions for using Preffi	87
4.5.1 Applying Preffi to one's own or other people's projects	87
4.5.2 Applying Preffi in various project stages and to various types of project	88
4.5.3 Using Preffi at specific or general level	89
4.5.4 Dealing with missing or unclear information	89
4.5.5 Function of Preffi	89
4.6 Conclusions	90

Chapter 5	Usefulness and reliability of a draft version of Preffi 2.0	91
5.1	Introduction	91
5.2	Methods	92
5.2.1	Study population	93
5.2.2	Measurement instruments	93
5.2.3	Data analysis	94
5.3	Results	96
5.3.1	Response	96
5.3.2	Project assessments	97
5.3.3	Reliability	101
5.3.4	Usefulness	105
5.3.5	Applying Preffi	107
5.4	Discussion and conclusions	108
5.4.1	Improving the instrument	110
5.4.2	Improving reliability	112
5.4.3	Expectations for the near future	114
Chapter 6	Project quality rating by experts and practitioners: experience with Preffi as a quality assessment instrument	115
6.1	Introduction	115
6.2	Methods	117
6.2.1	Study population	117
6.2.2	Project selection	117
6.2.3	Measurement instruments	118
6.2.4	Data analysis	118
6.3	Results	119
6.3.1	Reliability and accuracy of the intuitive assessments	119
6.3.2	Strictness of assessments: level of scores	120
6.3.3	Reliability of Preffi-based assessments compared to intuitive assessments	122
6.3.4	Reliability and accuracy of Preffi-based assessments	122
6.3.5	Notable aspects of project scoring	124
6.3.6	Consensus meeting	125
6.4	Discussion and conclusions	126
Chapter 7	Conclusions and future prospects	131
7.1	Introduction	131
7.2.	Lessons learned from the Preffi 1.0 programme in terms of the instrument's content, format and use	133
7.3	Preffi 2.0: development and validity	136
7.4	Preffi 2.0: usefulness and reliability	139

7.5 Overall conclusion on Preffi 2.0	142
7.6 Future prospects	143
7.6.1 Content	143
7.6.2 Format	144
7.6.3 Reliability	145
7.6.4 Role and status	146
References	149
Summary	165
Samenvatting	173
Curriculum vitae	181
Appendices	182
1. Preffi 1.0, scoring form	182
2. Calculating G and SEM in generalisability theory	184
3. Preffi 2.0 Scoring form (the foldout flap at the back of the book)	

Contents of the CD-ROM

Additional appendices

1. *Preffi 2.0 Scoring form*
2. *Preffi 2.0 Operationalisation and Norms*
3. *Preffi 2.0 User Manual*
4. *Preffi 2.0 Explanatory Guide*

The chapters and appendices included in the book are also on the CD-ROM.

De hoofdstukken en bijlagen zijn, voor zover beschikbaar, ook in het Nederlands opgenomen op de CD-ROM.

Preface

The completion of this thesis marks a major step in a process that started in 1994 with the project to design the first version of Preffi. The idea to write a PhD dissertation about this process arose some time ago. Gradually, my attention focused on the development and use of Preffi 2.0 as an instrument, and I have been working on this in an organised manner for the last two and a half years. Apart from a four-month sabbatical period from my busy job as a manager at NIGZ, I have had to do this in my spare time. Although working on this dissertation was sometimes a burden, it was also a pleasure, and I have learned a great deal from it.

My original background is that of a practitioner, though I have been a manager now for over twenty years. It is from this perspective that I have been following developments in health promotion. I tend to view research findings mostly from a practical perspective and to evaluate them for their practical usefulness. Being a manager, I greatly value rapid decisions and clear presentation and design, which sometimes clashes with the thoroughness that characterises the scientific effort. As a result, working on this dissertation meant a confrontation between the manager and the researcher in me, as well as between my management side and the views of my PhD supervisors. As I went along, I was able to experience the value of research and to position it alongside my practical and policy-making experience. This is highly relevant for my work as director of the NIGZ Centre for Knowledge and Quality Management, where we are always trying to synthesise knowledge obtained from researchers, policy-makers and practitioners.

Obviously, a programme like this is not a one-man effort, and I would like to express my great gratitude to some of the many people who have contributed to the present result. First and foremost among these, I have to mention my PhD supervisors. My close ties with Clemens Hosman date from as early as 1978. It was his enthusiasm for prevention that inspired me to start working in health promotion. This enthusiasm has not waned over the past 27 years, and during this time we have collaborated in many projects to promote the further development of our discipline. I regard the present dissertation as a fitting highlight of our collaboration. Gerjo Kok also provided hugely valuable contributions. It was he who stimulated me to express myself succinctly and clearly. In addition, he helped me put the work on this dissertation into perspective and provided vital psychological support.

Over the past ten years, the Preffi development programme has involved a great many colleagues, practitioners as well as researchers, who contributed in numer-

ous meetings, assessment rounds, advisory bodies and evaluation committees, for which I am very grateful to them.

I would like to specifically thank a number of persons who have played major parts in the process of designing Preffi over the past ten years. To begin with, there are my colleagues at NIGZ. Jolanda Keijsers was the instigator of the trend in Dutch health promotion which aims to synthesise knowledge obtained from researchers, policy-makers and practitioners, a trend that has resulted in the development of Preffi. It was Jolanda who asked me in 1994 to contribute to this process from the practitioner's perspective. To this day, she still devotes all her energy to this ideal. Hans Saan has provided major support, motivation and inspiration right from the start of the project. He has stimulated me by offering insights from the fields of quality management and project management, and, of course, he provided me with many creative ideas and solutions. The Preffi programme has gone through a series of stages, and in each of these I have had the privilege of working together with many wonderful colleagues in furthering the development of the instrument. Wilma van Driel collaborated with me on designing the first version of Preffi, while Margareth de Vries coordinated the implementation of Preffi 1.0. Leontien Hommels then completed the implementation programme and played an important part in designing Preffi 2.0. Machteld Ploeg was highly instrumental in defining Preffi's role and status within the quality management developments in Dutch health promotion. Louk Peters has been involved in the Preffi development programme since 1999, in his capacity as researcher and developer. Our many discussions, in which we tried to reconcile practical and research perspectives, have greatly contributed to Preffi 2.0 in its present form. He has co-authored most of the papers that I published on this subject.

Jacobine Hielkema, Theo Paulussen and Liesbeth van Meurs contributed to various components of the research project. Djoëke van Dale, Fee van der Poel and especially Henriëtte Kok have made valuable contributions to the further implementation of Preffi in recent years and still work with me on the Preffi team.

One of the aspects I have emphasised throughout this dissertation and in the Preffi programme is that content without context, and without practical support from institutional managers, can achieve very little. Hence, I owe a great debt of gratitude to Hans Hagendoorn, the director of LCG and later NIGZ, for his consistent personal support and appreciation of the Preffi project over the years. I am very happy to say that this has been continued by his successor, Hans Krosse, who has given me the opportunity and freedom to combine this PhD project with my job. I am also grateful to the NIGZ publishing department, for turning my dissertation into an attractive book.

NIGZ has invested a great deal of money and effort in the development and

implementation of Preffi. Major parts of the work were also co-funded by ZonMw (the Netherlands Organisation for Health Research and Development). Although it was not always easy to fit this expenditure into their regular research lines, the close involvement and support offered by Henk Smid, ZonMw's director, and Willem de Regt, its implementation adviser, have ensured that a number of Preffi development programmes could actually be carried out.

The research for the present dissertation involved quite a bit of complex statistical and methodological work, for which methodologist Paul Oosterveld was my helpful and expert coach. Jan Klerkx provided excellent and conscientious translations of my Dutch texts. I did much of the work for this dissertation at the Prevention Research Centre of the Clinical Psychology Department of Radboud University Nijmegen, where I have greatly appreciated the hospitality afforded by the colleagues at this institute.

Last, but of course by no means least, I have to express my immense gratitude to my 'home base', my family. My wife, Lilian Franse, who is the coordinator of a unit for mental health promotion, has been a wonderful sparring partner for me. Writing a dissertation requires a huge investment of time, energy and attention. It was great that Lilian, Jessie and Chris have allowed me the freedom to make that investment. Thank you very much.

Gerard Molleman

Chapter 1 Introduction

1.1 Background and problem definition

1.1.1. *Focusing on effectiveness*

The significance and value of health promotion are ultimately determined by its contribution to improving public health; good intentions are not enough. One of the main topics in the debate among health promotion workers in the last two decades has been the demand by society to assess the effectiveness and outcomes of programmes and interventions. Authorities at all levels, international as well as national and local, demand that health promotion specialists account for the effectiveness and outcomes of their interventions in increasingly clear and unambiguous terms. A new policy by the Dutch government to control the process from estimating policy costs to accounting for the amounts spent (the so-called VTBT operation) concentrates on the outcomes of efforts made (Ministerie van Financiën, 2002). At local level, audit offices are being created to assess the outcomes of local policies. The Association of Netherlands Municipalities (VNG) recommends that municipal authorities should implement their local health policy only with health promotion interventions that have been officially certified (VNG et al., 2005) and health insurers are only willing to pay for interventions that offer proven effectiveness. If no adequate account can be given of outcomes, it is expected that many subsidies will be withdrawn, something which has rarely happened so far.

This increased focus on outcome and effectiveness affects the entire health care system (Commissie Dunning, 1991). Effectiveness has been and still is a major component of the health care quality assurance programmes that have been started since the early 1990s.

It is against this background that the Dutch health promotion sector has also addressed the problem of effectiveness. The first clear signs of these were the publication of the book entitled *Effectiviteit van GVO* (effectiveness of health education) (Jonkers et al., 1988), the Wilhelmina Rouwenhorst lecture on effectiveness (NVGVO, 1988) and a conference in Nijmegen entitled 'Preventie in Zicht' (perspectives on prevention) on mental health promotion (Hosman et al., 1988). These showed that the question of effectiveness was highly relevant. Since then, there have been many studies into the efficacy and effectiveness of health promotion programmes (Bosma and Hosman, 1990; Kok and Borne, 1995; Paulussen

et al., 1997; Boddy, 1999; Keijsers and Vaandrager, 2000; Vaandrager and Keijsers, 2000; Rootman et al., 2002; Hosman et al., 2004)

These studies have shown that although health promotion programmes and interventions¹⁾ can work, their effectiveness varies greatly, ranging from highly effective, via moderately effective to poorly effective and ineffective. Also, programmes are not effective for everyone in all circumstances; the effects of an intervention are often limited to a particular segment of the target group in a particular situation and context, and cannot be generalised to other groups in other circumstances. In addition, when effective programmes are more widely implemented, they are often not applied with sufficient precision, as a result of which effective components lose some or all of their effect, comparable with the concept of treatment integrity. Meta-analyses tend to show low to intermediate average effect sizes ($ES = .25 - .30$).

The overall conclusion of the many studies would seem to be that there is still much effectiveness to be gained in terms of effects, reach and efficiency. Effects may relate to various levels of change, determinants, health and social outcomes, as well as to the efficacy of interventions and the effectiveness of interventions as implemented in everyday routine.

In short, there is a need to focus on effectiveness, a need which can be met in various ways.

1.1.2. Improving effectiveness

Various strategies can be used to improve the effectiveness of health promotion programmes and interventions in practice. Researchers, practitioners and support agencies may choose different strategies.

Contributions by researchers

In the last two decades, researchers have mainly focused on testing the efficacy and effectiveness of preventive interventions using rigorous guidelines. This is usually done in carefully controlled conditions, allowing clear proof to be obtained that an intervention works, provided it is implemented as intended. The ultimate aim of these research efforts is the large-scale dissemination and

1) This thesis uses the terms intervention, project and programme more or less interchangeably. The distinction between them is mainly one of size, comprehensiveness and complexity. The term intervention as used in this thesis thus often includes projects or programmes

implementation of only those interventions that are evidence-based and have been found to be effective, so they can be used as 'model programmes'. This approach greatly strengthens the position of health promotion and is therefore being widely applied in many countries.

Ideally, this should be a sound and well-structured approach. In actual practice, however, things are less simple, as the moderate effect sizes referred to above indicate. The large-scale implementation of interventions that have proved to be effective is often problematic and costly. Available funds may not be adequate. In addition, the developer of an intervention is often not responsible for its implementation. If an evidence-based programme whose effectiveness has been proven is widely disseminated, adopted and implemented, it usually needs to be adapted to the specific context in which it is to be applied (Greenberg et al., 2001; Grol and Wensink, 2001; Oldenburg and Parcel, 2002). In many cases, an intervention is actually only partly applied as intended – as is the case with many types of sound care programmes (Paulussen, 1994; Grol and Wensink, 2001; Dusenbury et al., 2004). The treatment integrity of interventions often leaves much to be desired. This makes it likely that the effects reported by the original research studies will not be achieved, or will be diminished, when the interventions are implemented elsewhere.

Another approach to increasing effectiveness is by focusing research on the determinants of the effects of programmes or interventions. This means trying to discover the reasons why the effects of interventions can vary, that is, what elements in the process of developing and implementing an intervention play a role in its ultimate effectiveness. These elements are also called effect predictors. There have been numerous empirical studies of individual programmes and interventions, as well as meta-analyses comparing large numbers of studies of interventions for their effect size. Studies have looked at specific elements, such as the use of interactive techniques or the presence or absence of a theoretical basis, in an attempt to explain differences in effect size (Llopis, 2002)

The knowledge thus obtained has been collected in more general form in textbooks of health promotion published by universities and other training institutions (Green and Kreuter, 1999; Woerkom and Meegeren, 1999; Brug et al., 2000; Bartholomew et al., 2001; Glanz et al., 2002; Green and Kreuter, 2004; Koelen and Ban, 2004). These textbooks use systematic planning models (such as the Precede – Proceed model or Intervention Mapping) for intervention development, and describe the successive steps in these models. They discuss existing theoretical models, their operational principles and the available empirical evidence supporting them. In a number of cases, the principles of effective

health promotion thus identified have also been translated into guidelines (Van den Broucke et al., 2004).

Another important aspect of research into determinants of effectiveness is that of analysing what practitioners actually do. Since they will be expected to start working in a more evidence-based fashion, it is important to assess their reasons for doing so or failing to do so. The current increased interest in implementation aspects has also led to greater focus on such analyses in relation to concrete interventions (Paulussen, 1994; Brug et al., 2000; Bartholomew et al., 2001; Fleuren et al., 2004).

There has so far been relatively little research into the methods used by the practitioners (Evans et al., 1994; Molleman and Nies, 1995; Peters et al., 2000; Jones and Donovan, 2004; Poel et al., 2005). This is nevertheless important, since the personal characteristics of practitioners and the circumstances in which they have to operate greatly affect the ultimate outcomes and the success of interventions.

In sum, the past 15 years have seen great progress in the creation of a sound scientific basis for health promotion, in terms of both theory development and empirical research.

Contributions by practitioners

Health promotion efforts in the Netherlands are largely the work of local health promotion specialists, who are employed by municipal health services (GGDs), mental health and addiction care institutes and home care agencies. When the development programme for the Preffi instrument was started, its local target group consisted of 650 professionals; by now, this has risen to nearly 1200 (Poel et al., 2005).

The practitioners' role is crucially important for the effectiveness of health promotion. This is certainly true for the Dutch situation, where it is the local practitioners who decide how projects are developed and implemented within the specific operational context.

Health promotion professionals tend to adapt programmes developed elsewhere and model programmes to their own context, since their superiors, usually the local authorities, want them to take rapid, concrete and visible action. They often have little time to develop interventions, as this is not regarded as their primary task. Nor do they have time to test the value or relevance of adaptations they make to existing interventions. At the same time, practitioners do

regularly develop new programmes and interventions for specific issues. Although a careful development and evaluation programme would be ideal in such cases, it would take years to obtain the necessary data. As a result, such a thorough approach is exceptional in practice. Of the 3000 projects which are currently being carried out in the Netherlands, only a small minority is probably sufficiently evidence-based (Poel, 2003).

The average practitioner finds it difficult to make optimal practical use of the multitude of – frequently equivocal – research findings on effectiveness and conditions for effectiveness. On average, current Dutch practitioners completed their basic training 8.6 years ago (Poel et al., 2005). Practitioners tend to read few scholarly journals, and the number of continuing education and refresher courses they attend is also limited. This means that recent research findings do not easily become widely disseminated among practitioners, a fact which is recognised by the practitioners themselves.

Attempts to promote more systematic practical knowledge in order to achieve greater effectiveness in health promotion are nearly always made in the context of projects set up jointly by practitioners and research and support agencies specialising in development and evaluation projects. A close relation between researchers and practitioners is essential for a sound analysis of the practical context and proposals for improvement within this context.

Contributions by a national support agency like NIGZ

The supporting role of an agency like the Netherlands Institute for Health Promotion and Disease Prevention (NIGZ) is primarily aimed at health promotion specialists. It is one of the tasks of NIGZ – as it was for its predecessor the Landelijk Centrum GVO (LCGVO) – to help bridge the gap between available research findings and their application by practitioners.

As a follow-up to its project on the effectiveness of health education (Jonkers et al., 1988), the LCGVO started a more comprehensive effectiveness programme in 1993, consisting of a number of components. The first component was an analysis of earlier research studies into interventions, with the aim of identifying conditions for the effectiveness of such interventions. To this end, the agency developed an instrument called *Analys* (Driel et al., 1994; Driel and Keijsers, 1997; Keijsers and Saan, 1998), which was then used for a whole series of reviews on the main topics in health promotion (reported in a series of 14 publications on the IUHPE project) (Veen et al., 1994).

The second element involved the development of a set of guidelines that should help health promotion practitioners improve the effectiveness of their actions. The guidelines were to be based on insights into effectiveness conditions derived from the reviews and research findings, as well as from knowledge and experience in practice. The set was based on the perspective of the intended users of the guidelines (the practitioners) and on their context, since the purpose was to find ways of systematically addressing the improvement of efficacy and effectiveness of interventions or programmes in practical operational processes. The type of knowledge that can help achieve this corresponds to the knowledge that is being collected and generated on determinants of the effectiveness of programmes in practice. This process of collecting and generating knowledge yields effect predictors that can be translated into principles and guidelines for practitioners. The process can assist practitioners in rapidly and adequately applying the latest research findings in order to work as effectively as possible, thus helping to reduce the gap between research and practice and improving the performance of the health promotion sector.

The above effectiveness programme covers not only the development of the set of guidelines but also its implementation and use. This involves raising awareness of the instrument and stimulating its use, for instance by offering training and support programmes and, where possible, promoting social pressure to get practitioners to actually use it.

1.1.3. International experiences and approaches

All countries that possess some health promotion infrastructure are working to improve the effectiveness of health promotion and disease prevention. The nature of this improvement process and the aspects of quality and effectiveness improvement that are addressed often depend on the available infrastructure and context in each country.

In some countries, like the US, it is the universities or research consortia that are primarily responsible for the development of prevention programmes. They tend to be mainly concerned with developing model programmes, which, after they have proved their value, are transferred to the practitioners for further implementation. A good example of this is the SAMSHA Model Programs National Dissemination System (SAMSHA and US Department of Health and Human Services, 2002). This is based on a national database of effective prevention programmes, all of which have undergone a review procedure and have been officially certified. Awareness of the database is being actively promoted,

and training programmes and technical support are available for the implementation of the programmes. Funds have been made available for this, and a monitoring instrument has been designed to measure the success of the SAMSHA system. Similar databases of model projects and summaries of reviews are also available at CDC in Atlanta, the Eppi centre and the Health Development Agency in Britain and the Cochrane Health Promotion and Public Health Field in Australia.

In many cases, the professional infrastructure for implementation and the political support for this infrastructure are beyond the responsibility of those who produce the databases of certified programmes. Most of the above processes are driven by researchers and have a clear top-down nature, and developers of model programmes often fail to take account of the wide range of practical circumstances in which such model programmes have to be applied. This may limit the usefulness of such model programmes. Gradually, the developers are beginning to understand this, and they are increasingly addressing the aspect of implementation and the problems encountered in practice. Various approaches are being used for this. The chairman of the Society of Prevention Research (SPR) recommends that programmes that have proved to be effective should be implemented as strictly as possible (programme fidelity), thereby emphasising once more the top-down nature of the implementation process (Botvin, 2004). The British HDA, like NIGZ and the Prevention Research Centre, prefer an approach in which developers are willing to learn from practitioners how the evidence found in research can be applied in practice (Hosman and Engels, 1999; Kelly et al., 2004). Countries where the health promotion profession is highly developed and where programme development takes place especially at a local level, often have some form of guideline development and quality improvement projects. Examples include Sweden (Berensson et al., 1996; Ovrerweit, 1996; Sanderson et al., 1996; Haglund et al., 1998; Ader et al., 2001) Switzerland (Health Switzerland, 2002), UK (Evans et al., 1994; Speller et al., 1998; French, 2004), Israel (Baron-Epel et al., 2004), Canada (CMHA and Health Canada, 1999; Cameron et al., 2001; Health Canada, 2001; Kahan and Goodstadt, 2001; Woodard, 2004), the United States (Work Group on Health Promotion and Community Development, 1995; US Department of Health and Human Services and CDC, 2001) and Australia (Public Health and Aged Community and Mental health, 2000; Australian Health Promotion Association, 2001; South East Health, 2003). Most of these guidelines take the form of questionnaires or checklists of key issues that professionals need to take into account when designing health promotion projects.

Depending on the context and the designer of the system, some countries also offer the option of entering existing projects in data systems and having them appraised in terms of quality, effectiveness and project management. This is done in Belgium (PK+)(Scheerder et al., 2003)(PK+), Germany, (Koch et al.), Australia (QIPPS) and the Netherlands (NIZW), as well as at EU level (IMPHA; Llopis). It is particularly the funding agencies that are in a position to demand that project teams make their knowledge and expertise available and allow it to be appraised.

An up-to-date survey of systems of quality-assessment tools, guidelines and databases is currently being completed jointly by the Flemish Institute for Health Promotion (VIG) and NIGZ, in the context of an EU project entitled 'Getting Evidence into Practice'(Vermeulen et al., 2005).

1.1.4. Conclusion

There is a need to improve the effectiveness of health promotion, and research has shown that there are indeed considerable opportunities for improvement. At the same time it is clear that there is a gap between what is known about effectiveness and what is actually applied in practice.

It is therefore important to invest in developing and expanding our understanding and the methods available to improve the quality and effectiveness of health promotion. In view of its mission, NIGZ decided to start a comprehensive effectiveness programme, a major component of which is the systematic development of a set of guidelines for practitioners. This set eventually became Preffi (Health Promotion Effectiveness Fostering Instrument), one of the instruments that can be used as part of a health promotion quality management policy. Before discussing Preffi in detail, we first discuss its role within quality management policy.

1.2 Role and status of the effectiveness programme in quality management

NIGZ regards effectiveness as a component of the broader concept of quality. The concept of quality refers to the degree to which a service meets previously defined requirements and expectations. Specifying the concept of quality is a process that should result in explicit criteria, requirements and expectations that must be met before something can be said to be of high quality. In his

review paper on quality criteria in the prevention, care, welfare and educational sectors, based on Dutch and international literature, Peters found that effectiveness is always a key aspect of quality (Peters and Keijzers, 2002).

However, the concept of quality includes not only effectiveness, but also output characteristics like relevance, reach, cost-effectiveness, ethical justification and the satisfaction or appreciation among professionals, clients and society (Donabedian, 1980; Walburg, 1997; Nabitz et al., 2000; Peters and Keijzers, 2002). If these output characteristics are to meet the required quality, certain requirements also need to be imposed on the input when designing and implementing a programme. It is the attention paid to both input and output that makes quality management a continuing and dynamic process, which also addresses the structure within which processes take place and the actual processes that are to produce the desired output. Each of these components can serve as a starting point for the attempts to improve the outcome.

Input characteristics are defined at two levels. The first is that of general contextual conditions for the development, implementation and evaluation of

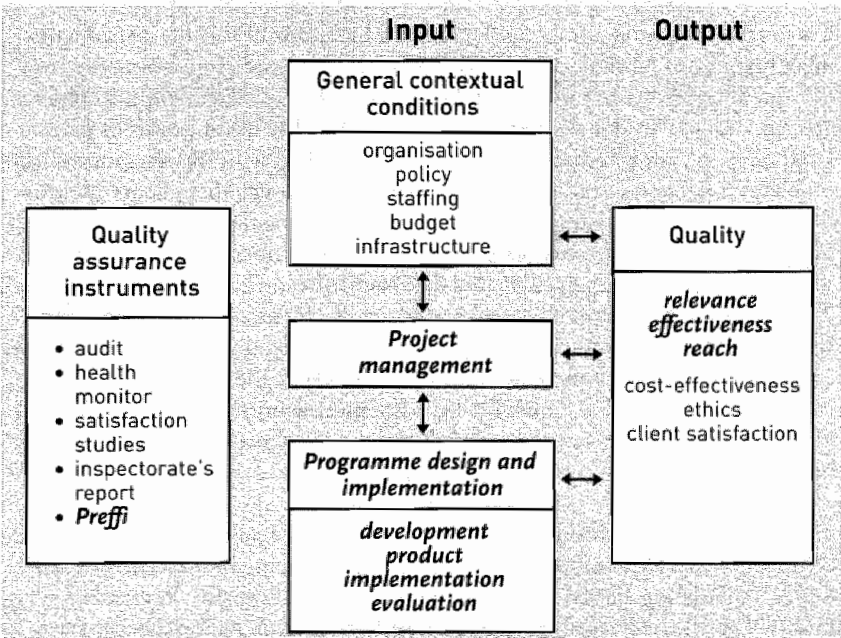


Figure 1. quality: input and output. Characteristics that are relevant to *Preffi* are shown in *italics and bold*.

health promotion programmes. Input characteristics at this level depend on aspects like infrastructure, the general policy of the agency developing a programme, the commissioning party, the characteristics of the setting in which the programme is to function, the nature of partnerships, staffing options, budgets, etc.

The second level is that of designing, implementing and evaluating the programme. The input characteristics at this level have to meet quality standards for programme development and implementation to allow the programme to produce an output that meets the quality requirements. Figure 1 illustrates this distinction.

In this figure, project management is shown as a mediating factor between the general contextual conditions and the design and implementation of a programme. Neither the programme nor the circumstances are static entities. They influence each other and go through certain changes while the programme is being developed and implemented. These changes affect the quality of the programme. For instance, establishing new partnerships often means that new types of intervention can be added to a programme.

The quality of a programme can be assessed by means of various instruments, including health monitors, audits, client satisfaction studies or reports by the inspectorate (see figure 1). Each of these instruments has its own possibilities and limitations. Health monitors are used to assess the health status of particular groups over time and are thus primarily concerned with the relevance of the problem. Satisfaction studies concentrate on assessing whether clients are satisfied, while audits can assess both the actual work being done and the operational procedures. Inspectorate's reports may be brief or comprehensive, and may include one or more of the above instruments.

The Dutch health care quality policy is based on the model developed by the European Foundation for Quality Management, called the EFQM model, and its Dutch translation, known as the INK model (after the Instituut Nederlandse Kwaliteit, a Dutch quality assurance institute) (Walburg, 1997; Nabitz et al., 2000). This model, which is shown in figure 2, also distinguishes between input and output characteristics.

The EFQM/INK model also provided the basic concepts and appearance of the certification scheme for health promotion which is being drafted by the the Dutch Foundation for the Harmonisation of External Quality Review in Health Care (HKZ) together with the health promotion sector. The first version of this certification scheme is expected to be completed by 2006.

The role of Preffi

Preffi is an instrument for quality improvement, to be used in designing and implementing programmes. Preffi focuses on the design and course of operational processes, which always need to be seen in the context within which a programme is being developed. Since none of the other quality assessment instruments used in health promotion assess the quality of operational processes, Preffi fills a gap in the quality assessment and improvement efforts.

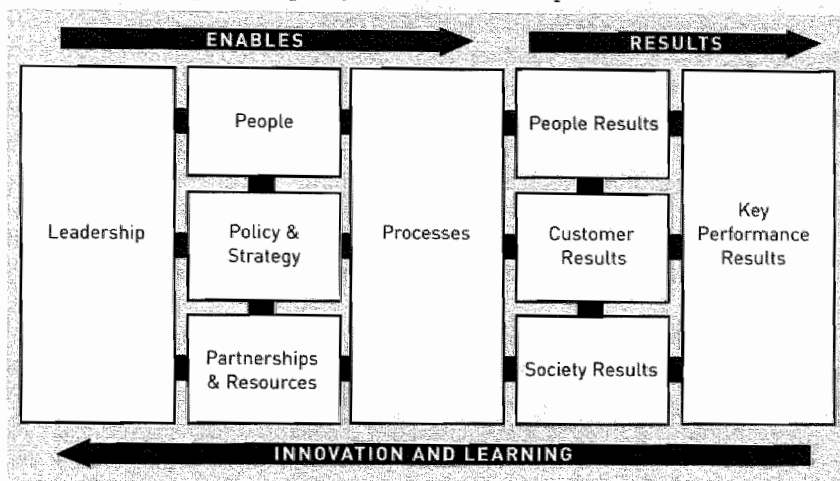


Figure 2: the EFQM/ INK model of quality management policy

Preffi's role can be further specified by clarifying the relation between Preffi and figures 1 and 2. Compared to the model in figure 1, Preffi does not include all output variables, concentrating instead on effectiveness, relevance and reach. This was a practical decision, rather than a matter of principle, and was based on capacity limitations and the view that not everything can be included at once in an instrument like Preffi. It is our intention to include aspects like cost-effectiveness and ethics more fully in later versions of Preffi. Actually, it should be noted that the output variables in Preffi are not independent: effectiveness, for instance, also requires a certain reach.

As far as the input side of figure 1 is concerned, Preffi tries to specify the quality issues relating to the panels headed 'design and implementation of the programme' and 'project management'. Although input variables relating to 'general contextual conditions' do impact on a project and its outcomes, we have not included them in the scope of Preffi, unless they have a direct effect on programme development. Project management includes ensuring that the pro-

programme content fits in with the prevailing contextual conditions. The input and output characteristics that are relevant to Preffi are shown in *italics* and **bold** in figure 1.

In terms of the EFQM/INK model in figure 2, Preffi concentrates on the processes, an aspect which is operationalised in detail in Preffi. Although the other components of the EFQM/INK model also occur in Preffi, they have not been specified in the same detail. The EFQM/INK components relating to staff, strategy and policy, resources and leadership are included in Preffi among the contextual conditions and project management elements. Some aspects of the EFQM/INK model's components relating to outcome (appreciation by staff, clients and society) are included in the Preffi criterion on evaluation; aspects relating to appreciation will affect the choices made for the operational processes.

1.3 Choices and requirements for a new instrument

Preffi was intended to assess and increase the quality of operational processes in order to improve the effectiveness of programmes and interventions. This led to the following questions we asked ourselves when developing the instrument:

- What defines effectiveness and which effect predictors affect effectiveness?
- What is Preffi's purpose within the wider context of quality management?
- In view of the available knowledge about guidelines and their implementation, what requirements should the instrument meet from a quality improvement perspective?
- What validity and reliability requirements should Preffi be able to meet as an assessment instrument?

1.3.1 The concept of effectiveness and the use of effect predictors

In developing Preffi, we decided to concentrate on the effectiveness of health promotion and the steps in the operational process that influence this effectiveness. Although all factors ultimately influence the effectiveness of an intervention, project or programme Preffi focuses on the theoretical steps that have to be taken and the choices that have to be made during the design and implementation of an intervention or programme. Project management aspects, including policy decisions by organisations and the allocation of resources, play a major role in the overall decision process. They are treated and referred to in this thesis as contextual conditions.

In developing Preffi 1.0 in 1995, it was decided to include those aspects or effect predictors that relate to the following dimensions (Molleman et al., 1995):

- the quality of the planning process;
- the attention paid to and the quality of the implementation;
- contextual conditions and project management.

A study of the literature and particularly the meta-analyses by Mullen and colleagues (Mullen, 1985; Kok et al., 1997) showed that the quality of the planning process was the main predictor of effectiveness.

Preffi 1.0 was based on the various planning stages of a project (Green and Kreuter, 1991; Kok, 1992). In addition, Preffi focuses on a number of principles of social learning, which we have included as 'effective techniques'. These are principles that can be included in virtually all interventions and that, as research has shown, can considerably increase the effectiveness of an intervention (Kok and Borne, 1995).

The practitioners' perspective was taken into account by including aspects of implementation, contextual conditions and project management. At that time (1995) these aspects had hardly been studied in the context of health promotion, although there was a broad consensus among practitioners, supporting institutions and researchers that these elements are important for the ultimate effect of an intervention (Saar, 1993; Paulussen, 1994; Kok and Borne, 1995). The Health Promotion Effectiveness Fostering Instrument we developed, called Preffi 1.0, is included in this thesis as appendix 1.

When it was decided to develop a new version, Preffi 2.0, NIGZ contacted the Prevention Research Centre (PRC) at Radboud University Nijmegen. We compared the dimensions underlying Preffi 1.0 with the dimensions of effect predictors used at PRC (Hosman and Llopis, 2002). This resulted in the identification of four dimensions that contribute to effectiveness: programme development, the programme itself, implementation and evaluation. These dimensions are illustrated by the model shown in figure 3.

The model is based on the structure of the planning process, and includes all dimensions used in Preffi 1.0. Within each dimension, developers need to make choices based on aspects of the programme itself (theoretical and empirical arguments) and aspects of project management and contextual conditions. This is shown in the model by the arrows going from the boxes representing these aspects to each of the dimensions.

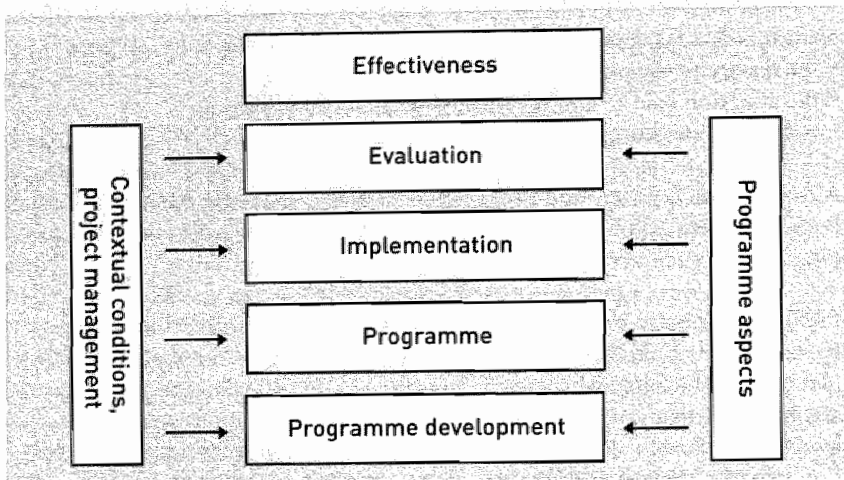


Figure 3: proposed model of dimensions of effect predictors

Preffi is intended to assess whether the elements of a programme have been selected and designed in such a way as to maximise the likelihood that the programme will be effective. To this end, Preffi includes **effect predictors**, that is, any type of criteria or variables that are demonstrably related to the programme's intended output (effectiveness). These are variables in the development, implementation and evaluation process that ensure that the choices made optimise the chances of an effective intervention. Chapter 4 of this thesis and the Explanatory Guide to Preffi (included as appendix on CD-ROM) discuss the arguments that determined our choice of criteria or effect predictors in Preffi 2.0

1.3.2 Preffi as a quality management instrument

Preffi intends to be a quality management instrument that can help increase the effectiveness of health promotion interventions and programmes by assessing their quality. To this end, it needs a set of criteria or effect predictors, as well as norms to indicate when a particular criterion has been met. This assessment can then be used to formulate proposals for improvement. After some time, the progress made can be reassessed and new proposals for improvement made. The instrument's assessment aspect should be subordinate to its intention to improve; one should be very cautious about using an instrument purely for quality assessment. A quality management instrument should help users understand

their own actions and should have a form which invites them to propose improvements. This requires that the users are actively involved in the process of developing the instrument.

Quality management instruments are dynamic entities, requiring continuous adjustment to the rapidly developing knowledge about effective health promotion programmes and their successful implementation. Regular updates of the instrument are required to incorporate the most recent research findings and experiences by practitioners. That is why we chose to add extensions like 1.0 and 2.0 to the instrument's name, allowing the requirements and specifications of the instrument to evolve with each new version.

We decided to develop an instrument that could basically be used to assess all types of health promotion and disease prevention projects and programmes. This meant that Preffi had to be worded in a kind of 'meta-language': it had to be composed of principles with a relatively general validity. Programme designers using Preffi should then consider how the Preffi principles can be applied to their own specific situation.

The consequence of our approach is that it is not possible to decide objectively, on the basis of an extensive survey of research literature, which would be 'the best intervention', regardless of the context. It is possible, however, to predict the likelihood of an intervention being effective in a particular situation, although the fact that an intervention works in one situation does not imply that it will also work in another situation. In the end, the effect will always depend on the circumstances. Preffi is about principles of efficacy and effectiveness. This means that, although using Preffi to assess a programme does not produce any absolute guarantees of effectiveness, such an assessment can be expected to increase the likelihood that the intended effectiveness will be achieved.

1.3.3 Requirements for Preffi based on knowledge about guidelines and their implementation

To maximise the chances of Preffi actually being used by health promotion specialists, we had to take the users' perspective into account right from the start of its development. The introduction and application of Preffi could be regarded as an innovation to its users, since no such instrument was available before 1995. Rogers listed the following characteristics that innovations should possess (Rogers, 1995):

- An innovation will only be adopted if it is found to offer advantages over the existing situation (relative advantage).

- An innovation should be appropriate for and consistent with the user's own values.
- Complexity is a major factor affecting the adopting of an innovation by users. What is decisive here is especially the perception of complexity (Assema et al., 1998). Since excessive complexity inhibits the adoption of an innovation, attempts should be made to keep the innovation as simple as possible and to pay a great deal of attention to clarity and accessibility.
- The step that a user must take in adopting an innovation should not be too big, and users should be allowed to try out the innovation first without definite commitment (tryability).
- It should be rapidly clear to users that adopting the innovation will produce results.

In a review of the literature on this topic, Grol e.a. listed a large number of characteristics of innovations that can facilitate or impede their actual application (Grol and Wensink, 2001). In addition to the factors already mentioned by Rogers, they included the following:

- the costs or investments required to apply the innovation;
- the degree of flexibility required to adapt the innovation to the needs or situation of the target group;
- an attractive design, which is not too long and focuses on essentials;
- the certainty that the innovation is useful (credibility);
- the degree to which the innovation concerns a key aspect of one's work and the impact it would have on the work;
- whether application of the innovation requires an individual or collective decision;
- the changes in organisational structure that would be required if the innovation is to be adopted.

Involving users in the development of innovations, in our case a set of guidelines, will influence its acceptability to practitioners. It creates support, allows the guidelines to be tailored to the needs, problems and work setting of those who have to use them and instils in their users a sense of ownership of the guidelines (Kotler and Roberto, 1991).

An important aspect of getting practitioners involved is that the guidelines should come from a source that is *familiar* to the intended users; research has shown, for instance, that 80-90% of medical specialists valued their own guidelines on a specific issue, while only 25% of the medical specialists valued the guidelines on the same issue that were produced by general practitioners (Grol and Wensink, 2001). Another important aspect is that the people involved in

the guideline development should be a *representative sample* of the intended group of users. The acceptability of the guidelines is also influenced by the *expertise* of those who develop and offer the instrument and those who become involved in this process. In terms of the Preffi project, this meant that users had to become involved in all stages of its development process. It was important to involve both pioneers and regular users from all sectors of health promotion (municipal health services, mental health and addiction care institutes, home care agencies and national health promotion agencies). In addition, researchers had to be involved to ensure acceptance by the scientific community.

Many of the aspects discussed above were investigated in numerous studies into the development and implementation of clinical guidelines (Grol and Wensink, 2001). These studies yielded valuable information on the requirements that guidelines have to meet, what they have to look like to maximise their use and which factors promote their use by practitioners. This knowledge has been synthesised in an international assessment instrument for the construction of guidelines, the AGREE instrument (Appraisal Instrument for guidelines, Research and Evaluation in Europe) (The Agree Collaboration, 2001; Burgers et al., 2003).

Some of the characteristics of guidelines that can contribute to their use by practitioners are (Burgers et al., 2001):

- concrete objectives;
- sound underpinning of recommendations by research findings;
- guidelines that are easily applicable, in that they fit in with the views and values of the target group;
- guidelines that are clearly and specifically worded;
- guidelines that can be flexibly applied and adapted to a wide range of settings;
- a clear structure and attractive layout.

Two types of objectives of guidelines can be distinguished:

1. They can be intended to provide professional support and improve the users' own practice.
2. They can be used for assessment, external inspection and possibly selection. These two options reflect the perspectives of quality improvement and quality assessment, respectively.

Whereas professionals prefer to use guidelines particularly for the first of these two objectives, that is, professional support, managers and sponsors may use guidelines to assess the quality of projects. Since the latter may have far-reach-

ing consequences for the professionals, unclear or unjustified expectations about guidelines may lead to resistance or to a fear of abuse, which may seriously hamper their use by the practitioners (Gevers and Biesaat, 1999). The two categories of objectives are not mutually exclusive, and may coincide. In such cases, however, it is important that users are clearly aware of this.

The lessons to be drawn from the development of guidelines were a useful reference for the development of Preffi, in terms of the appearance we wanted the instrument to have and the instructions for its use.

1.3.4 Requirements for Preffi in terms of validity and reliability

An instrument like Preffi needs to meet the usual requirements for such sets of guidelines, in that it has to be valid and reliable. Validity refers to whether an instrument actually measures what it is intended to measure. In the case of Preffi, what matters is effectiveness, or rather the likelihood of achieving the greatest possible effectiveness. Different forms of validity have been distinguished. The first form is content or construct validity, which involves determining what we mean by effectiveness and which conditions contribute to it. This construct validity relates to the criteria or items to be included in Preffi, as well as to their subsequent operationalisation and the norms devised for them. This meant that to develop Preffi, we had to consult regularly with researchers and practitioners to reach consensus about the factors that are relevant and the norms by which they are to be assessed.

The second form of validity is criterion validity that is, comparing the outcomes provided by an instrument with some external criterion. Another important form of validity is an instrument's predictive validity, which in our case means whether the Preffi scores can predict the effectiveness of a project. Predictive validity can only be assessed by means of longitudinal studies, examining the correlation between Preffi scores and the ultimate effects and outcomes of health promotion programmes or interventions after a number of years. This third form of validity was beyond the scope of the research reported on in this thesis.

In addition to validity, empirical research is needed to assess whether the intended users are able to apply the instrument and find it useful, and whether they are able to use Preffi to draw reliable conclusions.

Experience has shown that developing valid and reliable instruments for such complex matters as effectiveness is a slow, step-by-step process. The developers have to learn from the experience gained by its users. Its use must be monitored

and its reliability should preferably be ascertained by means of repeated empirical studies. Such empirical data can then be used to improve the instrument or the instructions for its use, in order to achieve the greatest possible validity and reliability.

The requirements that an instrument has to meet are also determined by the purpose that developers and users have in mind for it. In this respect, it is important to distinguish between the two possible objectives of an instrument like Preffi which were discussed above. If the instrument is to be a set of guidelines that allow practitioners to improve their methods, in other words, if the instrument is to be mostly diagnostic, the requirements in terms of content or construct validity are definitely important, although it is of course also relevant to know whether the assessments it produces are reliable.

If the instrument is to serve as an external instrument for inspection and assessment, the norms have to be stricter. In addition to its content validity and its criterion validity, the reliability of its assessments then becomes very important.

1.4 Developing and implementing Preffi (1994 – 2004)

This thesis focuses on Preffi 2.0, building on experiences with the first version, Preffi 1.0. The present section describes the activities that have taken place since the start of the project in 1994.

Preffi was developed by a project group at NIGZ (and its predecessor LCGVO) working on aspects of quality and professional development in health promotion. The group collaborated closely with experts from the research department and later from the Centre for Review and Implementation (CRIM). In 2003, the professional development group and CRIM merged with other departments to form the NIGZ-Centre for Knowledge and Quality Management. The Preffi programme was paid for by grants from the Preventiefonds and later from ZonMw (the Netherlands Organisation for Health Research and Development), as well as from NIGZ's own resources. Over the past decade, a total of close to one million euros has been invested in developing Preffi.

1.4.1 Principles

From the very beginning, it was our intention to develop an instrument that was meant to:

- bridge the gap between research findings and health promotion practice;
- achieve systematic improvements in the quality, and particularly the effectiveness, of health promotion programmes in practice, within prevailing conditions;
- stimulate the practitioners to 'learn how to learn', giving Preffi an added didactic function from the practitioners' perspective;
- allow an assessment and systematic evaluation of current practice.

As was described in section 1.1, the Preffi programme was initiated in 1994 by the Landelijk Centrum GVO (LCGVO) as a follow-up to the Analys programme (Driel et al., 1994). The Analys programme proposed and tested a systematic approach to reviewing the literature, the basic idea being that the results and conclusions of such systematic reviews could form a basis for developing a set of guidelines with which practitioners could improve the effectiveness of their interventions (Keijzers and Saan, 1998).

The development of such a set of guidelines started in 1994 and it was decided to design a quality management instrument from the perspective of the intended users, an instrument that would focus not only on assessing interventions and programmes but also on providing suggestions for their improvement.

1.4.2 Developing Preffi 1.0 (1994-1995)

The process of developing the set of guidelines known as Preffi 1.0 covered the period from April 1994 to August 1995. It started by analysing the information that had been systematically collected by the Analys programme. This yielded a number of basic assumptions and general principles about effectiveness and the conditions for achieving it, which could be incorporated in a set of guidelines. Basically, it was decided to work on the basis of a planning model like that developed by Green & Kreuter, supplemented by a number of specific characteristics and aspects of interventions, which had been shown by meta-analyses to be important contributors to the effectiveness of interventions (Kok et al., 1997). In the set of guidelines, we called these 'effective elements'. These included principles like providing room for a personal approach, giving feedback, using reward strategies, involving the social environment, etc.

We then consulted health promotion specialists and asked them what they considered to be crucial factors for the effectiveness of their interventions and programmes. They mentioned not only aspects of the programmes themselves, but particularly also the degree to which programmes are implemented, that is, their actual reach, as well as aspects of project management. They also empha-

sised the importance of the person who actually implements the programme, which we regarded as part of the project management.

Our search for information from research and practice yielded a seven-page list of aspects that were deemed to have an important influence on effectiveness.

We then asked the members of the Preffi advisory board, which included representatives of the various health promotion sectors in the Netherlands (municipal health services, mental health and addiction care institutes, home care agencies and national health promotion agencies) to apply this list to a particular project. This test was not a success. The participants unanimously commented that a set of guidelines would only be adopted in practice if it was short and simple. Thereupon we formulated a series of basic conditions that the set of guidelines had to meet:

- It should fit on one sheet of A4-size paper.
- It should not include more than ten items.
- It should ask users not only to assess programmes, but also invite them to improve their current practice.
- It should also include aspects of project management and implementation.

On the basis of this, we produced a draft version of Preffi 1.0, which was then discussed in expert meetings with representatives of the various sectors. In preparation for these meetings, the experts/practitioners were invited to apply the draft version to a project. The expert meetings were then used to assess whether the instrument was sufficiently useful and complete and whether its appearance was adequate. The suggestions made at the meetings were used to design the definitive Preffi 1.0 (included as appendix 1).

Preffi 1.0 was presented at a working conference, which also drew up recommendations for its large-scale implementation in the Netherlands. The conference recommended to continue the development of the instrument and to set up a support service for its development and implementation. It also recommended starting an implementation programme including a variety of activities that would appeal to various types of users in various stages of the process of using the Preffi (Molleman et al., 1995). The recommendations for the Preffi implementation programme were incorporated in a proposal. This was submitted to ZonMw, which agreed to fund the programme.

1.4.3 Implementing Preffi 1.0 (1997-2000)

Preffi 1.0 was published and made widely available in the autumn of 1995 (Molleman et al., 1995). Between that time and the start of the systematic

implementation programme in 1997, there were occasional lectures on Preffi and practitioners were occasionally given advice on its use. The systematic implementation was intended to promote the use of the instrument, and we also hoped to learn from the actual process of implementing Preffi.

The project plan for this implementation programme included various activities aimed at users, health promotion practitioners and strategic groups that can influence what professionals do, such as the Nederlandse Vereniging voor Preventie en Gezondheidsbevordering (Dutch association for prevention and health promotion, NVPG), managers of professional health promotion agencies, the umbrella organisations of municipal health services and mental health care (GGD NL and GGZ NL), supporting agencies like NIGZ and the Trimbos institute, as well as training and research institutes.

We studied the implementation process to assess the progress in the use of Preffi and identify its determinants. To this end, we did a baseline assessment (n=120) among professionals in February 1997 and a final assessment (n=316) in October 1999. The findings of the 1997 baseline assessment were used to formulate targets for the implementation programme, which were intended to be met by late 1999: 90% of the health promotion specialists had to be aware of Preffi's existence (compared to 75% in 1997), 40% had to have used Preffi at least once and 20% had to use Preffi as their standard instrument and had to be able to use it flexibly and efficiently. To achieve these targets, we used a series of interventions intended to:

- raise awareness of Preffi through mailings, publications, lectures and a Preffi award for best health promotion project;
- stimulate the use of Preffi by means of lectures, workshops and examples of the use of Preffi by colleagues;
- support the use of Preffi by providing a user manual, intake workshops, training sessions, intervision and consultancy.

Some of these interventions were general activities intended for all health promotion professionals, while others involved a more intensive type of support (intake, training and booster session) and were intended for a particularly motivated group of health promotion specialists (n=64), who had expressed an interest in applying Preffi in their own practice. The effects of this intensified programme were also assessed in our study (see chapter 3 of this thesis for the findings).

The implementation plan also included goals reflecting our intention that strategic target groups should stimulate the professionals to start using Preffi, for instance by incorporating the instrument in quality management systems

and by raising support for Preffi among researchers and trainers. In fact, however, very few activities were actually carried out among trainers and managers or those who decide on health promotion programmes. This was partly caused by a lack of capacity, and partly by the fact that we came to realise that influencing major stakeholders required a new strategy specifically tailored to this group, which would require additional effort.

In fact, the implementation programme was therefore limited to the immediate target group of the users. The tangible products of the implementation programme were a series of lectures, workshops and publications, a training package, a comprehensive handbook and a large number of examples of successful projects which came out of three rounds of Preffi awards (Molleman et al., 2000). The process of developing the Preffi handbook (Hommels et al., 2000) yielded a new format for Preffi 1.0, which already heralded the cyclic form of Preffi 2.0 (see figure 4). The Preffi 1.0 implementation programme ended in January 2000 with a working conference.

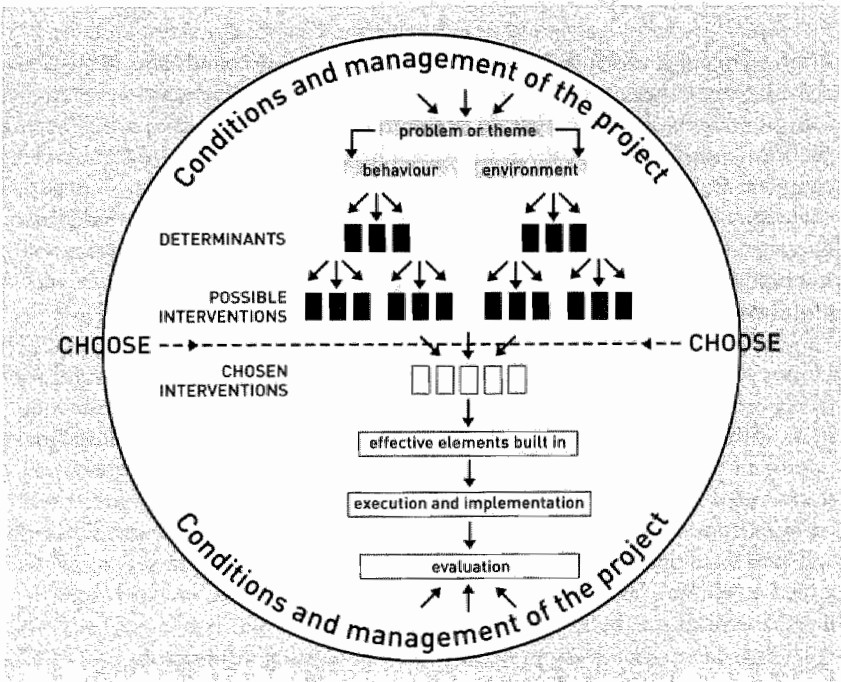


Figure 4: the Preffi 1.0 structure

1.4.4 Developing Preffi 2.0 (November 2000 – January 2003)

One of the lessons we learnt from the implementation programme was that there was a need for a new and updated version of Preffi, which would incorporate new research findings and the experiences of practitioners who had used Preffi 1.0. A systematic process to develop Preffi 2.0, supported financially by ZonMw, was started in November 2000. In this process, the NIGZ Preffi team collaborated with the Prevention Research Centre at Radboud University Nijmegen.

One major aim of the development programme was to raise Preffi's content or construct validity. This was attempted by means of a study of the literature on effect predictors, as well as by organising a critical dialogue with a group of representatives of the main approaches in Dutch health promotion and disease prevention research, which we called the *Wetenschappelijke Advies Raad* (Scientific advisory committee, WAR). The entire process of underpinning and operationalising Preffi 2.0 and developing norms for the scoring of its criteria was submitted to this WAR on several occasions.

After a while, we concluded that we also needed a group of advisors from among practitioners, which we called *Praktijk Advies Raad* (Practitioners' advisory committee, PAR), and in which about 50 practitioners took part. Whereas the WAR provided expert input to contribute to the construction of a sound, complete and valid instrument, the input by the PAR was particularly to help us design a practical, useful and clear form for the instrument. This underlined once again how important it is to develop such an instrument in constant consultation with its users, which had actually been our policy ever since the start of the Preffi programme in 1994.

During the first stage of the Preffi 2.0 development programme, we carefully re-examined the criteria/effect predictors of the first version, supplemented them with new insights and partly restructured them. The nature and designation of a number of criteria were altered, while criteria and clusters were restructured to make them more rational, and the total number of criteria was reduced from 49 to 39. All criteria were underpinned by new research findings, and they were operationalised and provided with norms to allow them to be scored as 'strong', 'moderate' or 'weak', indicating the degree to which a particular criterion is met by the project to be assessed.

A number of selection criteria were drawn up to allow us to select useful effect predictors for Preffi 2.0. These related to:

1. the available evidence for a relation between the effect predictor and the effectiveness of interventions,
2. the relevance of the effect predictor, in that the predictor had to contribute or could be expected to contribute to the effectiveness of interventions;
3. the generalisability of the effect predictor, in that the predictor had to be generally valid for all forms of health promotion;
4. the degree to which a predictor could be influenced or modified, since it would otherwise not be suitable for a quality improvement instrument;
5. the measurability of the effect predictor.

The draft set of effect predictors was evaluated by the project team and some members of the WAR. It turned out that selection criteria 3, 4 and 5 were met by virtually all proposed effect predictors. This left the relevance and the available evidence as selection criteria. We assigned a score between 1 and 5 to each effect predictor to indicate how much evidence was available that this predictor was related to effectiveness (see figure on page 4 in the Preffi 2.0 Explanatory Guide, provided on the CD-ROM).

The resulting draft version of Preffi 2.0 was tested in May – June 2002 by asking a group of 35 practitioners to assess a number of projects, thereby testing the instrument's usefulness and reliability. This test produced many useful data and suggestions for improvement (see chapter 5 of this thesis). After all suggestions had been incorporated, a definitive version of Preffi 2.0 was produced, in consultation with the WAR.

Preffi 2.0 consists of four components:

- a scoring form (see the foldout flap at the back of the book);
- a document called Operationalisations and Norms (provided on the CD-ROM);
- a User Manual (provided on the CD-ROM);
- a comprehensive Explanatory Guide, which explains the quality criteria (effect predictors) and discusses their importance for effectiveness (relevance) and the available evidence for their influence on effectiveness (evidence, including literature references) (provided on the CD-ROM).

These documents are included as appendices to this thesis, in cd-rom form, in Dutch and English.

The reliability and criterion validity of Preffi 2.0 were examined in the autumn of 2002 by asking three practitioners to assess 20 projects with the instrument. This assessment was compared with intuitive assessments of the same projects that the respondents had given beforehand by judging the main aspects of each

project without using Preffi. In addition, their assessment was compared with intuitive assessments of the main points of the same projects by three experts, in the expectation that the experts' assessments could be used as an external criterion (see chapter 6).

1.4.5 Implementing Preffi 2.0 (2003 – present)

After Preffi 2.0 had been launched at the third Dutch national conference on health promotion and disease prevention in January 2003, a new implementation programme was set up to introduce the new version to the practitioners. This programme included introductory workshops, as well as training sessions and support offered by the Preffi project team of the NIGZ Centre for Knowledge and Quality Management. We are also trying to get Preffi embedded in existing systems such as the QUI project bank and the ensuing in-depth studies, to develop local intervention portfolios around particular issues such as obesity and smoking, and to introduce Preffi in the new health promotion training programmes developed jointly by NIGZ and the Netherlands School of Public and Occupational Health (NSPOH). We strive to promote the adoption of Preffi's operational approach in the system of job specification criteria for health promotion professionals and the requirements that are going to be drawn up for the health promotion programme of the Dutch Foundation for the Harmonisation of External Quality Review in Health Care (HKZ).

1.4.6 International activities

In the course of the Preffi programme, many contacts were established with colleagues from other countries in Europe and beyond who are also working on guideline development. The Preffi team has given presentations and workshops on the instrument at all IUHPE conferences since the 1995 Turin conference (Herzyla 1997; Helsinki/Talinn 1999; Parijs 2001; Londen 2002; Parijs 2003; Melbourne 2004), as well as at some other international conferences on prevention in mental health care (Maastricht 1997; Atlanta 2000; Londen 2002; Auckland 2004). The Preffi 2.0 documents have meanwhile been translated into French and English, while Hungarian experts have already announced that they intend to have them translated and Norwegian practitioners are actively using Preffi 1.0.

The activities surrounding the introduction of Preffi have also led to a European project called 'Getting Evidence into Practice', in which NIGZ and its counter-

parts in other European countries collaborate to develop joint European guidelines for health promotion. The project is being coordinated by NIGZ.

1.5 General research questions and structure of this thesis

This thesis discusses the systematic development of Preffi 2.0. Based on the experience gained from the first 5 years of the Preffi 1.0 programme, it describes and justifies the new version and discusses the various studies we have undertaken to evaluate the new version's usefulness and reliability. The general research questions that this thesis tries to answer are:

1. What lessons can be learned from the Preffi 1.0 programme as regards the content, appearance and use of the instrument?
2. What was the process by which the new version Preffi 2.0 was developed and how were the validity requirements met?
3. Is Preffi 2.0 a sufficiently useful and reliable instrument?

Preffi 2.0 was developed by building on the knowledge and experience gained with Preffi 1.0 and its implementation. The first research question is discussed in chapter 2, which was originally published in the Dutch journal TSG.

Even if a sound instrument is available, this does not necessarily mean that it will be widely used. When developing Preffi 1.0, we tried to take into account whether the instrument could be implemented in Dutch health promotion practice. Implementing such an instrument requires a systematic approach, whose progress and results have to be monitored. If we were to improve the implementation process for Preffi, it was important to learn from the strong and weak points of the Preffi 1.0 implementation programme and to identify sources of resistance and opportunities. This process of learning how to learn from experience is the subject of chapter 3, which concentrates on the question to what degree Preffi 1.0 was implemented using a systematic approach, what factors influenced this process and what lessons can be learned for the future. This chapter was accepted for publication as an article in Health Education Research.

Chapter 4 describes the systematic development of Preffi 2.0. The definitive version was produced after a process of systematic consensus building with researchers and practitioners, in which its content or construct validity was assessed. The chapter outlines the definitive instrument that resulted from this process. The operationalisations and norms developed for Preffi 2.0, as well as

its user manual and the comprehensive explanatory guide, are included in this thesis as appendices. The explanatory guide justifies each criterion included in Preffi 2.0 and provides references to the available evidence underpinning each criterion. This chapter was originally published as an article in TSG in July 2003.¹ A modified version of the article was published in a special French edition of *Promotion & Education* (October 2004) and has been accepted for publication in the English edition of *Promotion & Education*, after a peer review procedure.

The draft version of Preffi 2.0 was tested for usefulness and reliability by a group of users. To this end, 35 health promotion workers each assessed 2 projects using this draft version. The study tried to answer 3 sets of questions:

1. How useful is Preffi 2.0; what do users think of the various components and aspects of the instrument; is this version an improvement over Preffi 1.0 and what is their opinion about the new version?
2. How could the draft version of Preffi 2.0 be further improved?
3. Can the reliability of the new Preffi version be determined?

This was intended as a pilot study that should allow us to assess the reliability more adequately in a further study. Its results are reported in chapter 5.

This study of the usefulness of the draft version yielded many suggestions for improvement, which were incorporated in the definitive version of Preffi 2.0. This definitive version was then used in a new study, in which projects were assessed with or without using the instrument, and in which the assessments obtained by practitioners using Preffi were compared with assessments of the same projects by experts. This study, which is described in chapter 6, first tested the assumption that the experts would agree in their assessments, so their judgements could be used as an external criterion. In addition, it compared the reliability and accuracy of the Preffi-based project assessments with those made without using Preffi. This study of the quality assessment of projects clearly revealed the complexity of such an assessment process in health promotion. This chapter has been submitted as an article to *Health Education Research*.

The final chapter presents the conclusions that can be drawn from the Preffi 2.0 development process and critically reviews the process. It tries to identify the current state of affairs for Preffi and the steps that can be taken to develop Preffi further in order to improve the quality of health promotion in the Netherlands.

Chapter 2 Preffi 1.0: development, experiences and starting points for a Preffi 2.0

2.1 Introduction

Since the early 1990s, health promotion specialists in the Netherlands have been trying to assess what determines the effectiveness of health promotion programmes and how projects can be designed in such a way as to maximise their effectiveness. The start of the prevention research programme by the Netherlands Organisation for Health Research and Development (ZonMw) has given effectiveness a prominent and permanent position on the agenda, and much research has been done over the past twenty years. This research has revealed large variations in the extent to which health promotion and disease prevention programmes are evidence-based (Gageldonk and Cuijpers, 1998; Hosman and Llopis, 1999; Vaandrager and Keijzers, 2000; Jansen et al., 2002). Although international research has provided evidence of the effectiveness of many programmes, the magnitude of such effects has often proved no more than moderate. Meta-analyses of large numbers of controlled effect studies (Durlak and Welsh, 1997; Brown et al., 2000; Llopis et al., submitted) have found average effect sizes (ES) of about 0.27, which means that participants in prevention programmes showed an average of 15% greater improvement than people in control groups. In addition, the effectiveness of such programmes was highly variable. While some programmes were highly effective, others were only moderately effective or had little or no effect, and programmes proved to be effective only among a specific part of the target group. Even effective programmes could be further improved, in terms of effect, reach and efficiency.

Developing, improving and implementing effective programmes require an understanding of evidence-based factors influencing the effectiveness of health promotion programmes. Such evidence-based factors are usually known in intervention research as effect predictors. Knowledge of effect predictors is highly valuable for programme developers as well as practitioners, provided it is translated into practical guidelines that allow them to systematically improve the effectiveness of their work, since practitioners are often required to design effective prevention programmes and interventions for a specific problem at short notice. By contrast, improving effectiveness through controlled research is

a slow and time-consuming process, involving long delays between programme development and the moments when information about effects is fed back. Although carefully designed and implemented effect studies are necessary, policy-makers and practitioners must also be able to rely on short-term feedback options. This means that it is important to have practical criteria that can help predict whether a particular programme or project can be expected to achieve the intended effects, even while the process of designing and implementing is still in progress.

This insight induced the Netherlands Institute for Health Promotion and Disease Prevention (NIGZ) in the mid-1990s to start the Preffi programme, a long-term project to improve the effectiveness of health promotion and disease prevention efforts by stimulating systematic, critical evaluation of programmes and projects. The core of the Preffi programme is the Preffi instrument, a set of guidelines comprising items relevant to the effectiveness of health promotion projects and programmes. These items reflect research findings on effect predictors, as well as insights into such predictors derived from critical discussions with practitioners.

Health promotion specialists are currently using Preffi mostly to help them develop and implement effective interventions. The instrument can also play a role in monitoring the progress of interventions and provides a framework for process evaluations of interventions.

An essential characteristic of the Preffi programme is that it tries to systematically improve and add to the set of instruments that can be used to increase the effectiveness of health promotion programmes. The Preffi team intends to produce regular updates of the instrument to further increase its validity and usefulness, incorporating new insights into effect predictors.

This chapter describes the experience gained with the use of Preffi over the period from 1995 to 2001, which has been systematically assessed among users at several moments in time. The article presents the findings of these assessments insofar as they are relevant to the second stage of the Preffi programme, which involves the development of the Preffi 2.0 version. By way of introduction, the next section discusses the format, content and role of Preffi 1.0.

2.2 Preffi

2.2.1 Background

In the mid-1980s, an interest in effectiveness issues arose among Dutch health promotion specialists, researchers and policy-makers (Hosman et al., 1988; Jonkers et al., 1988; Bosma and Hosman, 1990; Ministerie van Welzijn, Volksgezondheid en Cultuur, 1992). Effectiveness became one of the specified issues in the more general nationwide drive towards improving the quality of health care (including the so-called Leidschendam conferences). The quality assurance programme called *Actieprogramma kwaliteit voor gvo/preventie* (action programme on quality in health promotion) addressed (1) the development of a more professional attitude and the specialist profession of health promotion practitioner (in terms of a professional profile, code of conduct, accreditation and continuing education), (2) the organisations carrying out the actual health promotion and disease prevention activities and (3) the actual work, with effectiveness issues as the central theme (Saan, 1993).

In 1993, the predecessor of NIGZ (called Landelijk Centrum GVO) started a new programme with the aim of analysing and improving the effectiveness of health promotion interventions. This programme included the development of an instrument called *Analys*, which was able to analyse published reports on effect research in a standardised way (Driel et al., 1994). This instrument was developed in a joint effort with various health promotion researchers. *Analys* was tested on a range of thematic issues in health promotion and disease prevention settings, and was used in 1995 to review the international literature on the effectiveness of health promotion interventions in fifteen specified subject areas (Veen et al., 1994).

Although such reviews provided many valuable insights for practitioners, the problem was that most practitioners do not commonly consult reviews. What was lacking was a system to make the knowledge from these reviews systematically and readily available to practitioners and to make them use it to improve the quality of their work. This led to the idea of designing a set of guidelines for practitioners, based on research findings and systematic reviews.

2.2.2 Development

The process of developing such a set of guidelines was started in 1994. This process involved not only a survey of research insights and a general exploration

of guideline development, but also lengthy interviews with many practitioners (Green and Kreuter, 1991; Barnhorn and Walda, 1992; Damoiseaux et al., 1993; Grol, 1994; Janssen and Geelen, 1994; Tones and Tilford, 1994). The information thus collected was used to draw up draft guidelines, which were then discussed with the future users, that is, the health promotion specialists, in consultation rounds and workshops. In 1995, this resulted in the presentation of the first version of the Health Promotion Effectiveness Fostering Instrument, or Preffi 1.0 (Molleman et al., 1995).

2.2.3 Emphasising the practitioners' perspective

Our main concern in the process of designing Preffi was that it would have to be a useful instrument for practitioners. Hence, Preffi did not only include those effect predictors whose relevance had been adequately and convincingly demonstrated by empirical research, such as the quality of planning and the application of social learning principles (Kok and Borne, 1995). Because of the practical purpose, it also needed to include essential characteristics of implementation, management aspects and contextual conditions, such as feasibility, support and the adequate availability of staffing, money and time. Although little empirical knowledge was available on the relevance of these other aspects to effectiveness, experienced practitioners were able to provide useful insights in this respect.

Another aspect of 'usefulness' was the fact that practitioners often copy interventions developed elsewhere. The increased international exchange of evidence-based programmes and 'best practices', and the need to adapt these to specific local and cultural circumstances had led to a need to know which aspects have to be adapted and which ones have to remain unchanged to ensure that the intervention's effectiveness is not compromised (Hosman and Engels, 1999; Hosman, 2000). The new context in which such interventions are applied and implemented are rarely identical to that for which they were originally developed and in which they were found to be effective. This may have far-reaching consequences for the design of a project, and underlines the need for a set of guidelines providing general principles that can help design and implement operational processes in health promotion, so as to optimise the effectiveness of the resulting interventions. Practitioners would have to apply these effective elements in their own specific context, either in developing new interventions or in copying and adapting interventions that have proved their effectiveness elsewhere.

Our intention to make Preffi suitable for practitioners has greatly affected the instrument's design. The dialogue with potential users revealed that they were in need of a short, clearly structured list that would be easy to use and would stimulate them to assess the quality of their own products and operational processes. In addition, Preffi should not only allow them to assess quality, but particularly help to identify concrete points where improvements could be made. This has resulted in an instrument designed for both quality assessment and quality improvement.

2.2.4 Main focal points in Preffi

We designed Preffi in such a way that the effectiveness promoting conditions were reduced to ten focal points and a few contextual conditions, which we considered essential for the process of improving quality and effectiveness. We called these focal points 'criteria', and divided each criterion into a number of subcriteria. All of these were printed on one easily manageable A4-size card printed on both sides, which is the actual Preffi instrument.

The criteria were arranged in the order of the logical steps in a planning model:

- analysing the problem, including its nature and scale, as well as its determinants;
- making successive decisions about objectives, target groups and suitable intervention types, paying special attention to the inclusion of effective elements, as derived mostly from social learning theory;
- pre-testing and implementing the interventions;
- evaluating in terms of both process and effect;
- contextual conditions and project management, including the personal characteristics of the project manager.

Preffi invites its users to assess health promotion projects in terms of the various criteria, indicate points where improvements could be made, prioritise these points and implement them (see the appendix 1 showing Preffi 1.0).

2.2.5 Preffi from an international perspective

The process of developing Preffi fits in closely with insights obtained from similar quality assurance programmes in other countries, like Flanders (Van den Broucke et al., 1994), Sweden (Haglund et al., 1998; Ader et al., 2001), Britain (Speller et al., 1998), Germany (Christiansen, 1999), Switzerland (Switzerland, 2002) and Canada (Cameron et al., 2001; Kahan and Goodstadt, 2001). In

addition, researchers have recently taken various initiatives to develop new insights into factors influencing effectiveness (Green and Kreuter, 1999; Brug et al., 2000; Bartholomew et al., 2001). A new version of Preffi (called Preffi 2.0), which is being prepared, will incorporate the new insights gained from these programmes and studies. Preffi's emphasis on effectiveness issues, based on research findings as well as practical experience, makes it unique in the Dutch as well as international context.

2.3 A study of experiences with the use of Preffi to improve effectiveness

2.3.1 Questions addressed by the Preffi implementation programme

A systematic programme to implement Preffi 1.0 among Dutch health promotion practitioners, funded by ZonMw, started in 1997 (Molleman, 1999; Molleman et al., 2000; Peters et al., 2000). This implementation programme was supported by research, which particularly intended to answer the following questions:

- To what extent and how is Preffi 1.0 being used by health promotion specialists?
- What strengths and weaknesses of Preffi 1.0 have practitioners identified while using the instrument?
- What recommendations can be made for:
 - a. the content of a new Preffi version;
 - b. the further development of Preffi.

2.3.2 Research methods

In early 1997, a questionnaire was used for a baseline measurement (pre-test) among a representative sample of Dutch health promotion specialists. To this end, a quarter of all Dutch practitioners, selected at random, were approached. These 169 persons returned 120 completed questionnaires, that is, a representative response rate of 71%. A final measurement (post-test) was done in late 1999, for which half of all Dutch health promotion and disease prevention practitioners were approached. Of the 454 questionnaires sent out, 316 were returned in time (70%).

Our questionnaire on determinants of Preffi use was largely based on the Theory of Planned Behaviour and the ASE model, and included questions relating to

the respondents' awareness and use of Preffi, their opinions on Preffi, their self-effectiveness in applying the instrument, the support they had experienced from their social environment, their intention to use Preffi in the future and their preferences for support from the Preffi team (Vries et al., 1987; Ajzen, 1991). The 1999 post-test also included questions on the 'stage of change', to assess the extent to which Preffi was being applied (Prochaska et al., 1992). In addition to the above questionnaire assessments, we systematically collected data from 64 health promotion specialists who were highly motivated to use Preffi, during eight intake workshops held in 1998. The resulting data allowed us to evaluate how they actually applied the Preffi criteria in their own work routine. Further data were collected by means of process evaluations about the training programmes and consultative discussions provided by the Preffi team about using Preffi 1.0.

2.3.3 Interventions to promote the use of Preffi

A large number of implementation-oriented interventions were started, which aimed to:

- *raise awareness* of Preffi, for instance through publications, mailings, lectures and an annual 'Preffi award';
- *stimulate the use* of Preffi through lectures, workshops and articles discussing examples of the use of Preffi by colleagues;
- *support the use* of Preffi by providing a user manual, workshops, training sessions and consultancy.

This mixture of implementation interventions was based on the knowledge that the target group (the health promotion specialists) includes different subgroups in terms of their stage of adopting and using Preffi, viz., those unaware of Preffi, those doubting whether they want to use it, those learning to apply it and those intending to go on using it in their own professional situation. This subdivision is commonly used in implementation models (Paulussen, 1994; Grol and Wensink, 2001).

2.3.4 Results

The pre-test showed that many practitioners (75%) were already aware of the existence of Preffi at that time, as a result of publications on the development programme in 1994 and 1995, but that Preffi was as yet being applied only in isolated cases, rather than systematically. Based on this pre-test, we set the following goals for the Preffi implementation programme: after three years (i.e., at

the end of 1999) 90% of practitioners should be aware of the existence of Preffi, 40% should have used Preffi at least once, and 20% should be using Preffi as their standard instrument. In addition, we expected that the use of Preffi would be supported and stimulated by important others (institutional management, trainers, researchers, etc.).

The 1999 post-test showed that these goals had indeed been met: 96% of the practitioners were aware of the existence of Preffi, 50% were using the instrument occasionally and about 35 – 40% were using Preffi in accordance with its intended operationalisation. The percentage using it as intended was around 50% for practitioners in municipal health service and national health education institutions, around 40% for those working in mental healthcare (including the field of addiction) and around 25% for those working in home care.

Using the action and maintenance phase as a criterion, as advocated by Prochaska's transtheoretical model (Prochaska et al., 1992), resulted in a smaller percentage of practitioners using Preffi: 18%.

Preffi was generally being applied in a flexible manner, in the sense that users commonly applied only those parts of the instrument that were relevant to their specific context. Preffi was used before and during the project preparation stage, and to a slightly lesser extent during the project implementation stage or after its implementation.

The respondents reported that using Preffi stimulated them to check systematically whether and how they had addressed the various conditions for effective interventions. Our data analysis and the experiences reported during the training sessions produced the following overall picture.

Health promotion and disease prevention workers are very much action-oriented, and tend to make rapid choices of the intervention(s) to be used for a particular project. The decision which type of intervention to use is largely motivated by their earlier practical experience with intervention methods, their skills to implement certain interventions and their confidence in the effectiveness of such interventions, based on experience gained with them elsewhere. In addition, they tend to select intervention types that suit the tasks their organisation has set itself, that are acceptable to cooperating partners, that can be implemented within the available budget and that – as is often explicitly demanded by the institutional management or funding bodies – rapidly produce visible results.

We can conclude that the choice of a particular intervention is largely deter-

mined by practical arguments; the practitioners find it much more difficult to justify their choice of intervention on theoretical grounds. In most cases, no evidence is available that the intervention would be effective in the context in which it is to be applied, or if it is available, the practitioners may not be aware of it. Theoretical arguments for the decision to use a particular intervention are usually very general. Many practitioners find it difficult to apply theories to their own specific situation, and in fact they are rarely stimulated to do so. Many of them have no clear idea of determinants and the influence these have on the problem. Our analyses of the intake workshops (Peters et al., 2000) showed that 35% of the participants had never paid any attention to such determinants.

Objectives are usually worded in very general terms. Although 90% of them state what the intervention aims to achieve and for whom, only 29% indicate to what extent it is to be achieved and 39% state when it is to be achieved. According to the health promotion specialists, a measurable target was only identifiable in 40% of the cases. This makes it difficult to evaluate interventions and hence to learn from one's experiences. In only 34% of the cases was the project designed in consultation with the target group, and designers were hardly aware of the target group's willingness to change.

One of the goals of the Preffi implementation programme, getting important others (especially institutional management) to strongly support and stimulate the use of Preffi, was not achieved. The respondents indicated that they had only been stimulated to use it by their own colleagues and team, and hardly at all by their superiors or others.

The respondents generally had a positive attitude towards Preffi, and were especially positive about a number of aspects. They considered it important that the instrument provides an opportunity to analyse the effects of prevention projects, and that a common national set of criteria has been developed to promote further professional development. They regarded Preffi as an important, supportive and useful instrument. Our analysis of the post-test showed that the predictors of the intention to apply the instrument were the practitioners' attitudes towards Preffi, social influence, self-effectiveness and team embeddedness. Together it explained 46% of the variance in intention.

2.3.5 Advantages of using Preffi

The perceived importance of Preffi was also evident in practice, as we found during the training sessions and workshops. A systematic Preffi-based analysis

of the practitioners' own interventions yielded many suggestions for improvement and occasionally concrete adjustments to such interventions. Some of the more salient examples were the following:

- Although in many cases, the use of Preffi did not immediately lead to changes in the intervention itself, it improved the embedding of the intervention in a theoretical model and led to greater use of theory. This allowed the practitioners to become more aware of other important determinants and the corresponding perspectives on interventions. It also led to objectives, and chains or trees of objectives, becoming more theory-based.
- The practitioners became better able to motivate their objectives and to indicate which objectives were relevant to what part of the target group. The use of Preffi made the objectives more concrete, specific and measurable, and helped to avoid overambitious objectives. The question whether the intensity of a particular intervention would be sufficient to achieve the objective became more prominent, allowing a greater focus on evaluating projects.
- This procedure led to more realistic choices, allowing practitioners to estimate what can and cannot be achieved with a particular intervention within the specific contextual conditions. This can help prevent unrealistic expectations among institutional management and funding bodies, which are still all too common. More realistic ambitions should lead to greater satisfaction among health promotion specialists and their superiors.
- Since many practitioners were trained many years ago, working with Preffi may provide them with an opportunity to become familiar with new research findings and insights, especially as regards useful theories, evidence of the effectiveness of interventions, implementation and project management. In addition, they are stimulated to use reviews and other sources of information.

2.3.6 Disadvantages of using Preffi

When asked about the disadvantages of using Preffi in their daily routine, many practitioners mentioned that they were afraid it would take too much time. Interestingly, this was said especially by those who had not yet used the instrument. In addition, practical experience has shown that applying Preffi in its entirety is not easy. Although Preffi lists the criteria that are relevant to effectiveness, it does not specify what exactly needs to be done to satisfy each criterion. Many respondents mentioned that they would prefer to have a standard against which their own efforts could be tested. Some further aspects that were mentioned are discussed in more detail below.

2.4 Tension between preferred and actual situations

Working with Preffi often reveals contrasts between what is felt to be desirable and the practical opportunities to achieve this.

One such aspect is the desirability, which is now also recognised by practitioners, to work in an evidence-based fashion, that is, to use interventions that are aimed at proven risk factors and whose effectiveness has been established. In practical situations, however, there are numerous barriers to an evidence-based approach. In fact, it is often the context (i.e., opportunities, position, support, partnerships and experience) which decides what interventions are chosen. This shows that implementation aspects also constitute important preconditions for an evidence-based approach. Furthermore, effect studies often prove only that a particular intervention is effective in a specific context; in other words, 'evidence' is usually context-dependent. Whether an intervention will also be effective in one's own context can in practice only be predicted on the basis of insight into the active principles of the intervention and the minimum conditions for its effectiveness. This means that a new version of Preffi should allocate a more central role to 'context'.

The environment in which health promotion projects are implemented is highly complex. Causes are rarely unambiguous; instead, it is a matter of trying to find and apply the best possible solutions within the available opportunities and situational limitations. Since aspects like organisational conditions and support also affect the chances of success, the role and position of project managers are also important factors.

The importance of evaluating interventions and proving their effectiveness has received a great deal of attention in recent years. Much effort has been put into developing effect studies with maximum validity. While this has led to a significant increase in the number of research projects, it may also have caused too great an emphasis on controlled studies. As a result, many practitioners have the impression that they do not possess the necessary skills to carry out evaluations, the upshot of which is that not many evaluations are actually done. Other, easier types of evaluation, like process evaluation, case studies and monitoring systems, are also underused. This is regrettable, as generating feedback and learning from it is one of the most powerful incentives to improve one's methods. Our conclusion is that various elements of Preffi need to pay more explicit attention to monitoring, evaluation and types of evaluation.

Evidence is growing that participation of the target group in the development

of preventive policies is a crucial factor in reaching the target group, making projects succeed and achieving lasting effects (Vaandrager and Keyzers, 2000). The importance of target group participation is also being recognised elsewhere in society, for instance in social policies, the integration of migrants in society, urban development policies, crime prevention, etc. Whereas the importance and desirability of working from a community perspective have grown over the last few years, this approach has so far been rare in everyday health promotion and disease prevention work. Our workshop survey showed that practitioners collaborated with the target group in only 34% of cases (Peters et al., 2000). This means that a new version of Preffi should put greater emphasis on this type of partnership and on mustering support among the target group. An essential part of this is to take the cultural characteristics of the target group into account.

Not all practitioners have the skills required to apply such a 'bottom-up' approach. In addition, the relatively short lifespan of Dutch intervention and research projects also hampers the use and evaluation of a community approach, as this often requires sustained effort over a large number of years.

2.5 Consequences for Preffi 2.0

In 2001, the Preffi team started the process of developing a new version of Preffi, called Preffi 2.0. The aim of this process is to retain the valuable elements of the first version and to improve a number of other aspects. Some of the proposed improvements were already discussed above, while some others derive from practical experience with the use of Preffi 1.0 and the need to emphasise certain other aspects. Briefly, the changes relate to the following areas.

- *Content of the instrument*

New insights derived from research are to be incorporated. The content of the instrument is to be more fully specified as regards context and contextual conditions, the use of theories, types of evaluation, the importance of cultural aspects, participation by the target group and new insights regarding implementation strategies. In addition, all criteria listed in Preffi are to be operationalised and provided with norms (allowing projects to be assessed as strong, moderate or weak in terms of particular criteria).

- *Form of the instrument*

One of the attractive aspects of Preffi is that it fits on an A4-size card. Adding new elements to the instrument might make it unwieldy. It is therefore a

challenge to find a suitably compact form for the new version, one which also allows the scores attached to the operationalised criteria to be visualised.

The complexity of the real-world context does not always fit the linear structure of Preffi 1.0. A form needs to be found which does justice to the cyclical or iterative nature of many projects, making it possible to enter Preffi at different points.

An internet version of the instrument would make it possible to work with multiple layers, allowing users to select aspects they wish to explore in greater depth. This might involve calling up additional information from research findings or a list of alternatives available for specific choices to be made.

- *Role of Preffi*

The basic premise underlying Preffi, that of presenting a set of guidelines incorporating recent research findings and practical insights in order to maximise the effectiveness of health promotion projects, remains a strong concept. Hence, it is of great importance that we update and specify Preffi in an ongoing development process.

Preffi's original Dutch name, which literally meant Prevention Effectiveness Instrument, erroneously suggested that Preffi can provide objective and concrete measurements of the effectiveness of an intervention. Instead, Preffi is a quality assurance instrument, which can help identify the necessary conditions that need to be fulfilled for programmes to be effective, and hence to improve their effectiveness. Preffi's role in this respect needs to be clarified, and it has therefore been decided to rename the instrument Health Promotion Effect Management Instrument.

In addition, the instrument's usefulness can be enhanced by specifying that it can be applied within the framework of a larger quality assurance system, like the INK model (Instituut Nederlandse Kwaliteit, a Dutch quality assurance institute). It can thus become part of an audit system for health promotion.

2.6 Sustained implementation and development programme for Preffi

Preffi has become an accepted instrument to help achieve the necessary improvement of the quality of health promotion and disease prevention in the Netherlands. It is our expectation that its new version, Preffi 2.0, will be able to do this even more effectively. At the same time, however, the past few years have shown that improving the instrument itself is not enough. We also need to increase our efforts to implement it and provide support to its users. Implementing the use of a set of guidelines like Preffi is a long-term process.

Although during the six years of its development so far, Preffi's position in the field has been gradually strengthened, its implementation process is far from complete. There is still a need for periodic monitoring of the extent and nature of its implementation.

The Preffi implementation programme caters for a variety of target groups, including new practitioners who need to be made aware of the existence of Preffi, and experienced practitioners who are currently not using the instrument and need to be convinced of its usefulness and provided with support in applying it. We try to get the target groups to participate in the process of shaping and effecting this implementation process, in order to gradually create a sense of 'ownership' of Preffi among the professionals.

In addition to an extensive user manual and the provision of advice and training, many users would like to see an interactive Preffi website. Such a website, which is currently being developed, will also allow a more differentiated use of the instrument. It will be possible to develop different routes through Preffi for different types of project (e.g., community projects, mass-media lifestyle campaigns, implementing interventions that have proved to be effective elsewhere).

Consistent implementation of Preffi will remain a matter of perseverance. It is the practitioners themselves who will learn to use Preffi actively. Unfortunately, the institutional management of the organisations for which they work have so far failed to provide a supportive context of quality assurance policy and quality requirements, in which Preffi could be permanently incorporated. The use of the Preffi method also offers opportunities to generate systematic professional knowledge and so contribute to the further development of the profession. The next step in the Preffi development programme, the design of the new version Preffi 2.0, will be the subject of a later publication.

Chapter 3 Implementation of a quality assurance instrument (Preffi 1.0) to improve the effectiveness of health promotion in the Netherlands

3.1 Introduction

In the field of social and medical sciences, including health sciences, meta-analyses and reviews have found great differences in the extent to which programmes are effective and have revealed many factors influencing the effectiveness of programmes, usually called effect predictors (Contento et al., 1995; Durlak and Welsh, 1997; Kok et al., 1997; Tobler and Stratton, 1997; Hosman and Llopis, 1999; Brown et al., 2000; Cuijpers, 2000; McKay, 2000). New factors are constantly being explored and added. However, the available scientific knowledge about effectiveness and effect predictors is not utilized optimally in current practice: this is true for many sectors, including health promotion (HP) (Lomas, 1997; Shonkoff, 2000). HP specialists frequently have a minimum of time available to develop prevention programmes. It is therefore of great benefit to them if knowledge about effect predictors is translated into practical guidelines to systematically increase the effectiveness of their work. There are many opportunities to increase effectiveness, since a project's overall effectiveness is influenced by many factors. The major factors in the field of health promotion were included in Preffi 1.0, the first version of the Health Promotion Effectiveness Fostering Instrument (Molleman et al., 1995; Molleman, 1999).

Preffi is a set of guidelines for health promotion and prevention specialists, including 10 criteria with subcriteria that are regarded as essential for the quality and effectiveness of health promotion interventions. It was developed in 1994/1995 in close collaboration with HP researchers and practitioners.

The criteria follow the logical steps of a planning model: problem analysis, intervention development, implementation and evaluation. There is general consensus that systematic planning enhances effectiveness of interventions (Green and Kreuter, 1999; Bartholomew et al., 2001). In addition, Preffi stresses the importance of contextual conditions and project management. Preffi invites users to assess projects against the various criteria, indicate points

to be improved, prioritize them and achieve improvements. (See www.preffi.nl for more information about the criteria and subcriteria).

This article discusses an outcome evaluation study of implementing Preffi 1.0 in Dutch health promotion practice between 1997 and 2000. The outcomes pertain to practitioners' use of Preffi and psychosocial predictors of its use. Changes in effectiveness of programmes based on Preffi use were not tested since the latter is very difficult to assess and beyond the scope of this study. In the Netherlands, the health promotion infrastructure consists of about 900 highly trained HP specialists in public health and mental health.

3.2 Theoretical basis for the implementation process of Preffi

Innovations or new products do not simply sell themselves. Large-scale introduction requires specific action and consideration. Theoretical and empirical insights into implementation originate from social science, organizational theory and economics. While current models each have their own emphasis, they share many features. A successful diffusion process must take account of a number of crucial characteristics or principles: strategic planning, target group segmentation and an intervention mix (Paulussen, 1994; Rogers, 1995; King et al., 1998; Bartholomew et al., 2001; Grol and Wensink, 2001; Oldenburg and Parcel, 2002; Steckler et al., 2002; Grimshaw et al., 2004). Strategic planning involves a staged process to change the users' behaviour: dissemination (awareness of Preffi), adoption (deciding whether or not to use it), use or implementation and maintenance (continuing to use it, at both personal and institutional levels). Similar stages of behavioural change are also used in other models including the transtheoretical model (Prochaska et al., 2002). Stimulating the use of Preffi thus implies getting HP specialists to move on to the next stages of the diffusion process. And this calls for specific objectives and specific interventions for each stage.

The present study of the professionals' behaviour in using Preffi was based on the Theory of Planned Behaviour (Ajzen, 1988; Montano and Kasprzyk, 2002). The 'intention' (to start using Preffi) and the 'behaviour' (actually using it) in this model parallel the adoption and implementation stages in the above diffusion process. Attitude, social norm and self-efficacy in turn determine intention. These concepts from the Theory of Planned Behaviour partly overlap the determinants that affect the progress of the diffusion process, that is, the characteristics of the innovation itself, of the social and political context, of the

receiver and of the implementation strategy (Rogers, 1995). Perceived *characteristics of the innovation* include attitudinal beliefs, such as usefulness. The *social and political* context involves aspects like the organization employing the HP specialist and its existing culture, which may stimulate or impede the use of Preffi. Characteristics of the receiver include professional experience and training as well as aspects like self-efficacy in using Preffi. The *implementation strategy* for Preffi is the main independent variable in the present study, hypothesized to influence awareness, knowledge of content, attitude, social norms, self-efficacy, intention and use. Other independent variables include aspects of the receiver (training level, years of experience) and of the sociopolitical context (type of organization).

After Preffi 1.0 was developed in 1994/1995, a number of implementation activities were carried out in 1995 and 1996. In 1997 funding was obtained for a 3-year systematic implementation programme. At the start of that programme a nationwide pre-test among HP specialists was carried out to formulate objectives and develop interventions. The strategy included two approaches: general interventions and an intensified programme (see table 1). The intensified programme was developed in close collaboration with the target group.

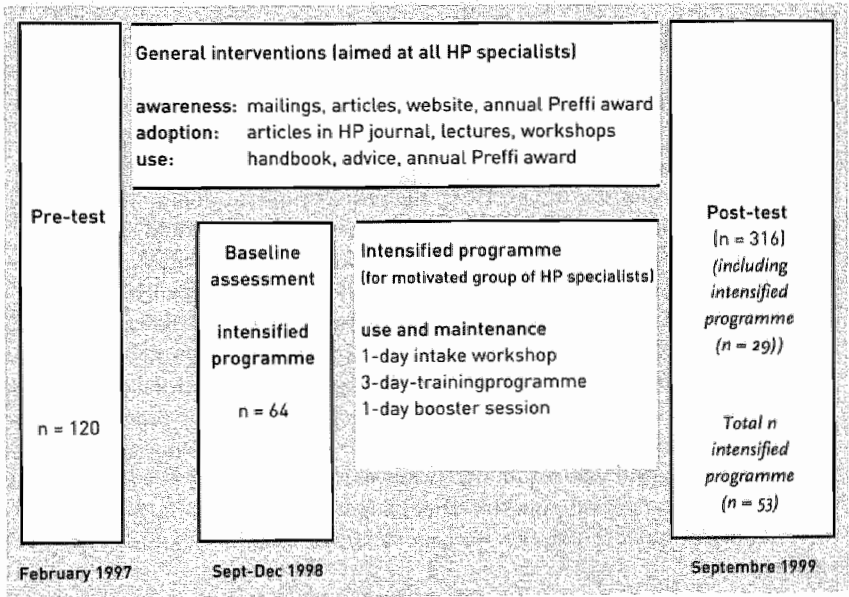


Table 1: overview of assessments and interventions

The following research questions were formulated:

1. To what extent have potential users progressed through the stages of the Preffi implementation process (dissemination, adoption and use) over the period 1997 - 1999?
2. Are there differences in the use of Preffi as a result of the general and the intensified intervention programme?
3. What factors explain the use of Preffi?

3.3 Research method

At national level, a pre-test was done in early 1997 and a post-test in late 1999, both using questionnaires, among two independent representative samples of potential users. In addition, data were gathered among participants of the intensified programme (see table 1).

Originally, the national study was planned as a cohort study, but names and addresses of pre-test respondents got lost due to the installation of new computers and staff turnover.

3.3.1 Samples

Nationwide assessments

For the pre-test a random sample of one quarter of all Dutch HP specialists (n=179) was drawn from a central register; 120 completed questionnaires were returned (67% response). In late 1999, the post-test questionnaire was sent to a random sample of half of all Dutch HP specialists (n=490). In addition, the post-test questionnaire was sent to all participants of the intensified programme not included in the representative sample. Of the 490 questionnaires sent out, 454 went to the representative sample. Of the latter, 316 were returned (70%).

The representativeness of the two independent samples for the total population of HP specialists was tested by comparing, on a group level, the pre-test data with the findings of a previous Trend Study into the characteristics of Dutch health promotion and prevention practitioners (Molleman and Nies, 1995) and by comparing pre-test and post-test data. Data were compared for type of organization, number of years of professional experience and type and level of training.

The training level of the Dutch professionals is high, with three-quarters having at least a university degree. One-third of the sample had a specialized HP/pre-

vention training. The average number of years of experience in this type of work is around 8.2 years. Almost 30% of those participating in the post-test had been working for three years or less and were not working in this profession at the time of the pre-test.

Compared to the 1994 Trend Study the pre-test sample had a higher average number of years of professional experience ($p < .001$). No other differences were found between the pre-test and the Trend Study and no differences were found between the pre-test and post-test samples.

Participants of the intensified programme

Data on most of the participants of the intensified programme were collected immediately before and 5-8 months after the programme. The pre-test for this group was in late 1998, while their post-test was combined with that of the general group. Data of 43 persons (67%) could be linked individually, as they participated in both the pre-test of the intensified programme and the general post-test. These 43 are not significantly different on any variable from the entire group who participated in the intensified programme.

Post-test data revealed that participants of the intensified programme had a significantly higher training level than non-participants ($p=0.04$). There were no significant differences between participants and non-participants for other background characteristics.

3.3.2 Instruments

National pre-test questionnaire

The pre-test questionnaire included three questions about background variables (type of organization; training; professional experience), a yes/no question on awareness of the existence of Preffi, and a question assessing the extent to which respondents were familiar with its content (not studied, superficially studied, studied some parts thoroughly, studied the entire content thoroughly). The use of Preffi was assessed by having respondents choose from a number of options (not yet used, derived some ideas from it, derived many ideas from it or using it as a general guideline, and fully mastered).

Psychosocial constructs were based on the theory of planned behaviour (Ajzen, 1991): attitudes, subjective norms, self-efficacy and intention. Items were drafted using the experiences of Preffi users, which were systematically recorded

during the development of Preffi (Molleman et al., 1995). All of these questions used five-point scales. The internal reliability of each scale was tested by calculating Cronbach's alpha. Table 2 specifies the psychometric properties of the scales, with pre-test information in the left-hand part (items marked 'pre-') and post-test information in the right-hand part.

At pre-test, attitudes about Preffi were assessed using 19 items. Factor analyses with varimax rotations showed that four useful subfactors were distinguishable within this attitude scale, together explaining 52.9% of the variance. See table 2 for more information.

National post-test questionnaire

The post-test questionnaire included all constructs from the pre-test questionnaire. In addition, new insights prompted us to include some additional concepts and questions in the 1999 post-test. An additional question about the use of Preffi assessed the stage of behavioural change (Prochaska et al., 1992). Also four new questions were included about the use of Preffi within the team, combined into a 'team embeddedness' scale. Three questions assessed actual social support from the respondents' immediate environment ($\alpha=.81$), whereas at pre-test expected support was measured. In addition, scales assessing attitudes, self-efficacy and intention were expanded with several questions. See table 2 for more information. Unlike the pre-test questionnaire, the post-test used ten-point scales for the questions; this difference was due to staff turnover.

Attitudes about Preffi were assessed by 23 items ($\alpha=.81$). Factor analyses with varimax rotations yielded four useful subfactors, which together explained 48.5% of the variance. The pre-test and post-test attitude factors did not include exactly the same items, but roughly corresponded, as pre-test factors 1, 2 and 3 resembled post-test factors 2, 1 and 4, respectively. In view of the large number of missing values for attitude items, these analyses were conducted only for those who reported having used Preffi (based on the question about stage of change, $n=130$). To reduce the number of missing values and increase the power of the various scales, we used mean substitution as a data imputation method. The number of imputed scores per scale did not exceed 25% of the total number of items on the scale.

Specific pre-test questionnaire for the intensified programme

Before the start of the intensified programme, we used a baseline assessment to collect various data. Background data, familiarity with the content, self-efficacy and social influence were assessed with the same questions used in the pre-test of the nationwide study. The use of Preffi was assessed by stage of change ques-

tions, as in the post-test. Intention was assessed by two questions (use in next six months, next project; $\alpha = .74$). Attitude was assessed by four general attitude items also included in the pre-test ($\alpha=.65$). See the items marked 'inten' in the left-hand part of table 2.

3.3.3 Data analysis

Data from the two independent samples in the pre-test and post-test were compared at group level. Data from the pre-test of the intensified programme were linked to post-test data at individual level, and analysed as such.

In the post-test many respondents who had not yet worked with Preffi did not answer questions about attitudes, social norms, self-efficacy and intention. This explains the smaller number of respondents included in the post-test analysis. A probable cause was that the wording of these items unfortunately may have led non-Preffi-using respondents to consider these items not applicable to them (e.g., "I use Preffi because ..."). Missing values were as high as 50% for some items. The post-test data thus refer largely to those who reported having used Preffi in some way.

Pre-test and post-test scores were compared by converting the pre-test data into ten-point scales, by changing the pre-test scores 1-5 into scores 2, 3.5, 5.5, 7.5 and 9. We used $p<0.05$ as the significance level in two-sided tests. Stepwise regression analyses were used to identify predictors of the intention, and simultaneous logistical regression analyses for predictors of the various stages of implementation. Background characteristics (training, experience and organization) were converted into dummy variables to enable us to include them in the regression and logistical regression analyses. The cut-off point for the number of years of professional experience was three years. Since the number of institutions varied considerably per sector, effect coding was used for the type of organization.

3.4 Results

3.4.1 Familiarity with and use of Preffi

Differences between pre-test and post-test

As table 3 shows, significantly more respondents had heard of Preffi at the post-test compared to the pre-test (96% vs. 76%), but respondents had not grown

Scales and items within these scales:		pre-test	inten. prog.	post-test
	← expected	α = .84	α = .63 inten.	α = .81
	immediate colleagues			
	immediate superior management	pre- pre-	inten. inten.	
	team embeddedness			α = .67
	use Prefri in team-meetings			
	Prefri fits well in our quality-system used Prefri last year at least once in team-meetings last year my organisation recognized Prefri as a quality instrument			
	self-efficacy	pre-	inten.	α = .84
	capable in using Prefri in next project			
	can assess when to use which part of Prefri can assess with which intensity use parts of Prefri			
	intention	α = .82	α = .74 inten. inten.	α = .73
	intend to use Prefri at start next project	pre-		
	intend to start (or keep) using Prefri as a team we will use Prefri in at least 1 project for future project Prefri will be quality-instrument	pre-		

attitude scales: system. = systematic approach; useful = useful instrument; recogn. = external recognition; clarity = clarity of instrument; negat. = negative aspects
pre = item included in pre-test questionnaire
inten. = item included in pre-test questionnaire intensive program

Table 2: scales and their psychometric properties

more familiar with its content. As for the type of use of Preffi, the number of HP specialists in the entire group who had derived some or many ideas from it or were using it as their general guideline had significantly increased, from 42% to 50%. This includes the 30% of HP specialists who had been working as such for three years or less at post-test and were probably not aware of Preffi at the start of their careers.

Influence of participation in the intensified programme

The findings in table 3 also show that the significant differences between the pre-test and post-test scores can almost entirely be attributed to participants of the intensified programme. Among non-participants there are no significant differences between pre-test and post-test in the extent to which Preffi was used.

There were significant differences between participants of the intensified programme and non-participants in the degree of familiarity with the content of Preffi and in the extent to which they derived some or many ideas from it. The post-test also assessed Preffi use in terms of stages of change. In 1999, 44% of the HP specialists were using Preffi, with 18% situated in the action and maintenance stages. Again, there were differences between those who did or did not attend the intensified programme (39% vs 83%, $\chi^2 = 23.8$ df=3 $p = .000$).

Trying Preffi had not induced any of the respondents to refuse to use it in the future, that is, there was no relapse.

The right side of table 3 presents the scores for participants of the intensified programme prior to and after participation. The table shows a favourable shift for the levels of familiarity ($p = .004$) and stage of change ($p = .015$).

3.4.2 Attitude, support, self-efficacy and intention

Pre-test scores of all respondents (not in table) on attitudes, support, self-efficacy and intention were generally slightly favourable, ranging from 5.7 (attitude factor 4 'clarity of instrument') to 6.4 (intention) on a scale ranging from 1-10. An exception was attitude factor 1 'systematic approach' with a score of 7.6.

Differences between pre-test and post-test

Since at post-test it was mostly those respondents using Preffi who answered the determinant items, we only compared the pre-test and post-test scores of users (see upper part of table 4). Pre-test scores of users did not differ significantly from the entire pre-test sample.

	Pretest - Posttest			Post-test with/without intensified programme				Before/after intensified programme			
	pre	post	p.	post no prog	p. related to pre-test	post progr	p. related to prog +/-	before programme	after programme	p.	
	(n=120)	(n=316)		(n=287)	to pre-test	(n=29)		(n=43)	(n=43)		
	%	%		%		%		%	%		
aware of Prefri	76	96	.000	96	.000	100	n.s.				
level of familiarity			n.s.		n.s.		.000			.004	
not studied	34	28		31		0		9	0		
superficially	28	32		34		17		49	12	5	
parts thoroughly	20	22		19		44		28	32	14	
thoroughly	18	19		17		38		14	56	24	
level of use			.017		n.s.		.000				
not yet used	58	50		54		10					
derived some ideas	26	21		20		31					
many ideas, general guideline	16	29		26		59					
stage of change				(n=270)			.000			.015	
precontemplation		32		36		0		5	0		
contemplation		24		25		17		51	12	5	
preparation		26		22		55		30	42	18	
action + maintenance		18		17		28		14	46	20	
relapse		0		0		0					

Table 3 : pre-post scores on awareness, familiarity and use for general group and intensified programme

Post-test scores of Preffi users on attitude items were slightly more favourable and the difference was significant for the sub factors 'systematic approach' and 'useful instrument'. The intention to use Preffi had also increased significantly, while the change in self-efficacy narrowly failed to reach significance ($p=.052$). As for social support, expectations at pre-test were moderately favourable, but post-test scores revealed hardly any actual social support. Immediate superiors and especially management were hardly reported to have stimulated Preffi use. The post-test scores were almost the same for respondents who participated in the intensified programme and those who didn't, with the exception of self-efficacy. Participants of the intensified programme gave a significantly higher score of 7.8 ($p=.021$) than in the pre-test.

Differences in scores before and after the intensified programme

Participation in the intensified programme produced sharp and significant increases in self-efficacy and intention (see bottom part of table 4). Actual social influence was significantly less than expected, except for immediate colleagues.

Attitudes showed a significant reduction after the intensified programme but remained favourable or at least neutral. Participants who had used Preffi had recognized that not all projects are suitable for Preffi analysis. Scores on the value of analysing effects of intervention programmes had also fallen. Participants of the intensified programme assessed Preffi as less concrete than non-participants ($p=.038$, not in table).

3.4.3 Predictors of intention and use

We performed regression analyses to predict intentions and use from psychosocial and background characteristics. Because psychosocial items in the post-test questionnaire were answered mainly by Preffi users, prediction of intention was analyzed only for the pre-test sample. As table 5 shows, all psychosocial variables correlated significantly with intentions; this was much less the case for background characteristics. Due to correlations between potential predictor variables not all these variables had a unique contribution to the prediction of intention. Multiple regression analysis revealed that intention was explained for 60%, with contributions from social influence, the attitude factor 'systematic approach', self-efficacy and the background variable professional experience.

comparison pre-test vs post-test	scores of Preffi users only				
	pre-test (n=50)		post-test (n=158)		p.
	x	sd	x	sd	
attitude scales:					
systematic approach (f1)	7.7	.73	8.0	1.08	.021
useful instrument (f2)	6.5	.73	6.8	1.00	.013
external recognition (f3)	6.2	1.07	6.3	1.44	.640
clarity of instrument (f4)	5.8	1.27	6.0	1.51	.358
self-efficacy	6.9	1.60	7.4	1.84	.052
	expected social influence		actual social influence		
social influence	6.1	1.42	4.2	2.03	.000
immediate colleagues	6.0	1.67	5.0	2.42	.001
immediate superior	6.3	1.69	4.7	2.66	.000
management	6.1	1.63	3.3	2.14	.000
intention	6.9	1.26	7.4	1.58	.028
intensified training programme (n=43)	before programme		after programme		p.
	x	sd	x	sd	
attitude	7.6	.68	7.0	1.2	.014
important to analyse effect	8.4	.93	7.7	1.6	.024
Preffi suitable for all projects	6.9	.96	5.2	2.7	.000
Preffi useful instrument to analyse effects	7.0	1.2	7.2	1.8	.669
Good to have 1 national list to analyse interventions	8.1	1.0	8.1	1.6	.835
	expected social influence		actual social influence		
social influence	7.0	1.8	5.1	1.0	.000
immediate colleagues	6.7	1.2	6.4	1.9	.371
immediate superior	7.5	1.3	5.3	2.5	.000
management	7.0	1.5	3.7	2.2	.000
self-efficacy	5.3	2.1	8.2	1.3	.000
intention	7.0	.92	7.8	1.2	.001
I intend to start or keep using Preffi	6.7	1.1	8.1	1.3	.000
I intend to use Preffi at start next project	7.2	1.0	7.4	1.7	.607

Table 4: pre-post scores on psychosocial constructs for general group and intensified programme

pre-test scales	r	β
att.f1: systematic approach	.60***	.304 ***
att.f2: useful instrument	.54***	n.s.
att.f3: external recognition	.41***	n.s.
att.f4: clarity of instrument	.29**	n.s.
social influence	.62***	.395 ***
self-efficacy	.55***	.288 ***
background characteristics		
type of organization	n.s.	n.s.
professional experience	.08	.177 **
educational level	.08	n.s.
specialized HP training	.11	n.s.
multiple R		R ² = .60
*p<.05, **p<.01 *** p<.001		

Table 5: prediction of intention at pre-test from psychosocial and background characteristics: zero-order correlations (r) and β weights

Predicting use and the stages of change

Logistic regression analyses of the stages of change were carried out for the post-test sample of users. These analyses yielded no significant finding when comparing contemplation on the one hand and preparation, use and maintenance on the other. This makes sense, as nearly all data refer to users, who have moved beyond contemplation.

However, use and maintenance of Preffi, compared to preparation, was explained by more positive scores on the attitude factor 'instrument' and by greater self-efficacy and team embeddedness. The distinction of preparation/use versus maintenance was solely explained by team embeddedness.

3.5 Discussion

Before interpreting our findings we first want to comment on the methodological limitations of our study.

One major point is that the data acquired in this study are based on two successive, independent samples from the population of health promotion special-

	prep - action/ maintenance	prep/action - maintenance
post-test scales	O.R.	O.R.
att.f1: useful instrument	1,989*	1,842
att.f2: systematic approach	0,942	0,839
att.f3: negative aspects	0,895	0,932
att.f4: external recognition	0,706	0,771
social influence	0,862	0,919
self-efficacy	1,328 *	1,240
team embeddedness	1,646 **	1,521 **
background characteristics		
type of organization	n.s	n.s
professional experience	1,69	1,723
educational level	1,224	0,951
specialized HP training	0,693	0,636
	$\chi^2 = (14) 32,432$ $p = .003 (n=126)$ $C\&S R^2 = .227$	$\chi^2 = (14) 27,090$ $p = .019 (n=126)$ $C\&S R^2 = .193$
*p<.05, **p<.01 *** p<.001		

Table 6: prediction of stages of change at post-test from psychosocial and background characteristics

ists, rather than two successive assessments in the same cohort. It is, however, reasonable to assume that both samples were representative of the total population. The only exception was in the larger than average number of years of professional experience in the pre-test sample. Since professional experience is a predictor of higher intention to use Preffi in the pre-test sample, the pre-test findings may be more favourable than the real situation. This means that the estimation of effects is conservative.

Although emphatically invited to fill in all questions in the questionnaire, a large group of respondents who had not yet worked with Preffi failed to complete many of the questions about determinants at post-test. As said before, this is probably due to the wording of items. The data presented here relate mostly to users.

Since the five-point scales used at pre-test were converted to ten-point scales to allow comparisons with post-test data, this comparison should be interpreted with some caution. Hence we chose a conversion factor that takes this into account.

Our findings relate to self-reported use, and do not reflect whether Preffi is actually being used as intended. It seems reasonable to assume (see also Paulussen, 1994) that it will only to some extent be used as intended. Data published elsewhere confirm this, showing that Preffi users thought they were adequately applying the Preffi principles, even though this was only partially true (Molleman, 1999).

The present study shows that a general, nationwide implementation strategy can yield a high level of familiarity and a positive intention and attitude about using an instrument like Preffi. It does not, however, achieve an increase in actual use. We conclude that these general activities can achieve a basic level of awareness. Getting people to actually use Preffi requires more effort. Attending a few workshops or lectures on Preffi, as was done in our general strategy, was found not to be enough. By contrast, participating in the intensified programme did yield a significant increase in the actual use of Preffi. These findings are in agreement with those of other implementation studies (Brug et al., 2000; Grol and Wensink, 2001).

A striking finding of our study is the lack of support for working with Preffi among the management of the organizations employing the HP specialists. Managers seem to regard working with Preffi as the HP specialists' own professional responsibility. In view of the major effect of social support and team embeddedness we found, such support could have had a very favourable influence. Future implementation efforts should pay more attention to embedding innovations in organizations and teams.

The pre-test measure we used to assess the use of Preffi (none, little, considerable, as a general guideline) was obviously not clear or precise enough, as it hardly revealed the influence of the various determinants. The stages of change model provided a better definition of the level of use. The logistic regression analysis of the post-test data provided useful indications of the determinants that were relevant in the various stages of the process. Team embeddedness, which is a stimulating environment, self-efficacy and perceived usefulness of the instrument were found to be important in the transition from preparation to action/maintenance, as was to be expected. The transition from preparation/action to maintenance was mainly affected by team embeddedness, under-

lining again the importance of institutionalisation for permanent implementation. As for attitudes, items about the value of a systematic approach influenced intentions but not actual use; use was predicted by items about practical usefulness, which makes sense.

Our findings support the view that an implementation process consists of a number of stages, each demanding its own approach in the implementation strategy, as was shown by the regression analyses. The research model used, based on the Theory of Planned Behaviour and the stages of change concept, provided a useful frame of reference to structure and analyse this type of innovation process and revealed important determinants.

The positive attitude towards Preffi slightly eroded with actual use. This may seem surprising but has been reported before (Paulussen, 1994; Brug et al., 2000), as the unrealistic optimism people may have felt before they actually tried it was dampened and reduced to realistic proportions.

3.6 Practical implications

The implementation process has revealed the value and necessity of developing such processes jointly with the target group. New interventions must be developed and applied in an ongoing process, in order to provide the target group with a sense of ownership of Preffi.

Implementation demands great effort and sufficient capacity, and is easily underestimated (Grol and Wensink, 2001; Oldenburg and Parcel, 2002). Introducing a quality assurance instrument like Preffi requires a long-term investment, which can only be achieved by appointing a change agent who is constantly engaged in and feels responsible for the implementation. This aspect is often overlooked in ideas about implementation. Its consequences are far-reaching, as the supervisor must be able to continuously invest in the implementation process. This may clash with funding practices, in which institutions may largely depend on short-term project grants; this is true for the current Dutch situation.

The Preffi implementation process should continue to consist of a mixture of interventions, with complementing general and intensified approaches. More emphasis will have to be put on the intensified programme, which allows close collaboration with the user group.

Our study has revealed the importance of team embeddedness, whereas stimulation by management of the use of Preffi among health promotion specialists was found to be negligible. Hence, improvements can be expected from the use of an implementation strategy aimed at the management and the HP specialists' immediate superiors. Such a strategy should concentrate on their views and motives to improve the effectiveness of their organizations. Recently, Preffi was included in the quality assessment procedure of the Dutch national Health Inspectorate, which may add to its external recognition among management. Also, the Preffi Award has been extended with an award for the manager who best supports use of the instrument in his or her team or organisation.

In addition to these implementation activities, the Preffi instrument itself will need to be further improved. Recently a new, updated and more refined version of Preffi was developed, based on increased knowledge about effectiveness and needs assessment among users in training courses, who asked for operationalization and norms (Molleman et al., 2003; Peters et al., 2003). It is expected that this more precisely defined Preffi instrument can be more easily applied and that this will contribute to its implementation rate.

Chapter 4 Preffi 2.0: systematic development of a quality assurance instrument for health promotion

4.1 Introduction

The second version of the Health Promotion Effect management Instrument, or Preffi 2.0, was launched in January 2003. Preffi 2.0 is a quality assurance instrument that allows users to assess whether a health promotion project has been designed in such a way as to maximise the chances of it being effective. This also allows them to identify points that can be improved and to take action to achieve this improvement.

Preffi 2.0 was developed by a joint project group including experts from the Netherlands Institute for Health Promotion and Disease Prevention (NIGZ) and Radboud University Nijmegen, and builds on the experience gained with the instrument's first version (Preffi 1.0) over the past six years (Molleman et al., 1995; Molleman, 1999; Molleman et al., 2000; Peters et al., 2000; Molleman and Hosman, 2003). This experience had shown that four aspects of Preffi needed improvement: its content, the norms, its format and its role (Molleman and Hosman, 2003). In terms of content, the aim was to incorporate into the second version all recent insights provided by health promotion research and practice. The Preffi criteria had to be operationalised in such a way as to allow users to assess their own projects against a normative standard and to allow third parties to assess a project. The format developed for the new Preffi version had to do justice to the cyclical and iterative nature of many health promotion projects and had to allow users to use its components in the order that suited them best. Finally, it was deemed necessary to clarify Preffi's role and status as a quality assurance instrument aimed at identifying and improving conditions for the effectiveness of projects. The instrument's original Dutch name, which literally translates as Prevention Effectiveness Instrument incorrectly suggested that Preffi actually measured the effectiveness of projects. Finally, defining the role and status of the new version meant that its scientific legitimacy was enhanced.

This article describes the systematic development of Preffi 2.0, explains the instrument and reports findings of a study into its usefulness and the resulting suggestions for the practical application of Preffi.

4.2 Developing Preffi 2.0

In developing Preffi 2.0, much effort was invested in providing a solid scientific basis and in operationalising all Preffi criteria. To achieve these aims, we collaborated closely with the Prevention Research Centre at Radboud University Nijmegen and a scientific advisory committee (Wetenschappelijke Advies Raad, WAR), consisting of five experts representing the main approaches to health promotion research (see Note a). In addition, a practitioners' advisory committee (Praktijk Advies Raad, PAR), composed of 53 Dutch health promotion specialists (see Note b), ensured that the users' perspective would not be neglected.

The development process of Preffi 2.0 was based on a model consisting of a number of systematic steps, and used various methods of formative evaluation, viz., product-oriented, expert-oriented and target group-oriented methods (Jong and Schellens, 2000). In product-oriented methods, it is the designers themselves who indicate the aspects of the draft design on which they base their evaluation, while in the other two methods, the design is submitted to external experts and to members of the target group, respectively. The Preffi 2.0 design process started with product-oriented methods (steps 1-5), followed by expert-oriented methods (steps 4-6) and target group-oriented methods (steps 6-7).

- Step 1.* The Preffi project team first tried to define the instrument's role and status within the general context of quality management (see also figure 1 in chapter 1). As is common in this context, a distinction was made between input and output (Donabedian, 1980; Walburg, 1997). The project team decided that, as regards output, Preffi should focus on the effectiveness, relevance and reach of programmes, rather than on other output characteristics such as cost-effectiveness or client satisfaction. On the input side, Preffi's focus was to be on the actual operational processes involved in the design and implementation of programmes, rather than on structural aspects of organisations (infrastructure, institutional policies, staffing, etc.). At the same time, these aspects cannot be completely separated from the actual operational processes, and they were included in Preffi as contextual conditions for these processes.
- Step 2.* After we had decided on the role and status of Preffi, we searched for a general structuring principle for the actual operational processes. The principle we chose was that of a systematic approach, since there is broad consensus that a systematic approach improves the effectiveness

of prevention programmes (Mullen, 1985; Kok, 1992; Glanz et al., 1997; Green and Kreuter, 1999; Brug et al., 2000; Bartholomew et al., 2001). This meant that the content of Preffi 2.0 did not differ fundamentally from that of Preffi 1.0. We distinguished five dimensions: the contextual conditions dimension and the four stages in a systematic approach (analysis, development, implementation and evaluation). In each of these dimensions, a set of effect predictors could be identified that would be relevant to that particular dimension. The main effect predictors of each of these five dimensions were explicitised.

- Step 3.* In order to stimulate the debate on the selection of effect predictors, we first defined five selection criteria, viz., relevance, scientific evidence, generalisability, modifiability and measurability. These criteria were then used by the members of the project team to evaluate the effect predictors included in Preffi 1.0. We found that virtually all effect predictors in Preffi 1.0 met the criteria of 'generalisability' and 'modifiability', which meant that these criteria had hardly any discriminatory value for the selection process. Nor did 'measurability' discriminate, as it was met by all predictors, in terms of making predictors as clear and measurable as possible for the target group of practitioners.

This left the criteria of 'relevance' and 'evidence'. The project team decided that the relevance of an effect predictor for the effectiveness of a programme was the most important criterion. Relevance refers to the proven or assumed impact of particular characteristics of a project or of other conditions on its effectiveness.

Preffi is more than just a checklist of effect predictors: it should also enable users to understand the relevance of these predictors. The scores we gave for scientific evidence reflect the instrument's 'evidence-based' character. A broad definition of the concept of 'evidence' was used, for one thing because the predictive value of some effect predictors that are deemed important is hard to prove in randomised controlled trials, which is usually regarded as the highest form of evidence (Cochrane Collaboration; CSAP (Center for Substance Abuse Prevention) ; NHS Centre for Reviews and Dissemination (CRD); Rada et al., 1999). The principles of a systematic approach, for instance, are usually based on rational argument, consensus and the findings of repeated case studies. There is still no such thing as an internationally accepted hierarchy of types of evidence (McQueen, 2000; McQueen and Anderson, 2002), and support is growing for the view that other methods of verification

than controlled experiments are also legitimate (Tones, 1999; Koelen et al., 2001; McQueen and Anderson, 2002).

- Step 4.* The results of the application of these selection criteria to Preffi 1.0 were first discussed within the project team, then with the individual members of the scientific advisory committee WAR and then in a plenary meeting with the entire committee. The outcome of these consultations was used to define a number of Preffi 1.0 predictors more precisely, though no effect predictors were removed. The consultations also covered ideas for the addition of new effect predictors and for adjusting existing effect predictors. These ideas were based on new insights derived from health promotion research and practice, as well as on the experience gained with Preffi 1.0. This is discussed in more detail below.
- Step 5.* After the project team and the WAR had agreed on the main aspects of selecting the effect predictors, these were specified on the basis of research findings from the literature and the expertise available within the project team and the WAR. This was done first of all in an Explanatory Guide that described the nature of each predictor, its relation with effectiveness (relevance) and the available evidence for this relation. A second document provided a further operationalisation of each predictor in terms of specific questions, as well as the corresponding norms, based on argued consensus between the project team and the WAR. This should help the intended users, that is, the practitioners, decide to what extent their projects pay adequate attention to each predictor, while also proposing specific aspects that need to be improved.
- Step 6.* In addition to the WAR, the practitioners' advisory committee, or PAR, was also involved in the various consultations. The WAR and PAR were asked to give their opinion on the Explanatory Guide and operationalisation document for Preffi 2.0. Their feedback was used to introduce various adjustments, such as subdividing effect predictors, clarifying explanations and adjusting the operationalisations and norms.
- Step 7.* After we had completed a draft version of Preffi 2.0 (consisting of a scoring form, the Explanatory Guide, the operationalisation document and a user manual), it was tested for practicability among 35 experienced practitioners from a range of institutes. This group, which also

included 21 PAR members, was asked to use the draft version of Preffi 2.0 to assess two project descriptions, and then to complete a questionnaire (including both open and closed questions) about their experiences while applying the instrument and their opinions about the various elements of the draft version. Supplementary interviews were held with 10 of the practitioners. The findings of this pilot study were then used to adjust certain aspects of the content and appearance of the draft version in the definitive version of Preffi 2.0, and to revise ideas about the use of the instrument. Since the findings of this study and the changes they led to cannot be understood without knowing the definitive version of Preffi 2.0, the next section first describes this definitive version. We then discuss some findings of the study and the lessons we have drawn from them for the use of Preffi 2.0.

4.3 Preffi 2.0

Preffi 2.0 consists of four components (Molleman et al., 2003; Peters et al., 2003):

- A Scoring Form (the foldout flap at the back of the book): this is a four-page document with a list of quality criteria (i.e., effect predictors), scoring categories for these criteria, a box to note points to be improved and the Preffi 2.0 model, a visual representation of the main principles behind Preffi.
- The Operationalisation and Norms document (provided on the CD-ROM): this is a 22-page document that provides operationalisations for each of the quality criteria using one or more yes/no-type questions and norms (categories of scores), based on the answers to the operationalisation questions.
- The Explanatory Guide (provided on the CD-ROM) : this is an extensive (73-page) document which provides further details on the quality criteria (effect predictors) and discusses their importance for the assessment of effectiveness (relevance) and the available evidence for their impact on effectiveness (including literature references).
- The User Manual (provided on the CD-ROM): this is a 7-page document which explains Preffi 2.0 and provides instructions for the use of the instrument and each of its components.

The next subsections discuss the basic concepts behind Preffi 2.0, the main content changes relative to Preffi 1.0, and the scoring method. Later sections discuss the instructions for use.

4.3.1 The Preffi 2.0 model

The main conceptual elements are represented in the Preffi 2.0 model (see figure 1), which emphasises the dynamic nature of health promotion projects, that is, the permanent interaction between content and contextual conditions and the cyclical nature of the process of health promotion.

The central part of the model shows the steps involved in the systematic design and implementation of a project, that is, the actual process of health promotion: analysing the problem, choosing and designing the right intervention, implementing it and evaluating it. These process steps are shown in lozenge shapes because each step first involves looking at a wide range of options (divergence) and then choosing from among these options on the basis of content aspects and contextual conditions (convergence). To give an example: the analysis of the problem ideally involves identifying all potential causes/determinants, after which a choice is made of determinants to be addressed in the intervention, based on substantive arguments, relevance, modifiability and practicability.

The lozenges overlap because the selection process should always take the consequences and options for the next stage into consideration. For instance, in choosing a particular intervention, designers should already take into account the opportunities for implementation. The evaluation relates to all moments in the process where choices have to be made, which is why arrows point to all these moments.

The effectiveness of interventions and the choices that can be made within those interventions are partly determined by contextual conditions like the support for the project, the capacity available for its implementation and the quality of the leadership provided by the project manager.

The arrows surrounding the lozenge shapes indicate the moments when choices have to be made. The arrow pointing at the choice of an intervention is larger, to indicate that this is the point where the influence of contextual conditions is often particularly strong.

4.3.2 The list of criteria and changes relative to the previous version

The 39 quality criteria (effect predictors) included in Preffi 2.0 have been subdivided into eight clusters (see the foldout flap at the back of the book). The

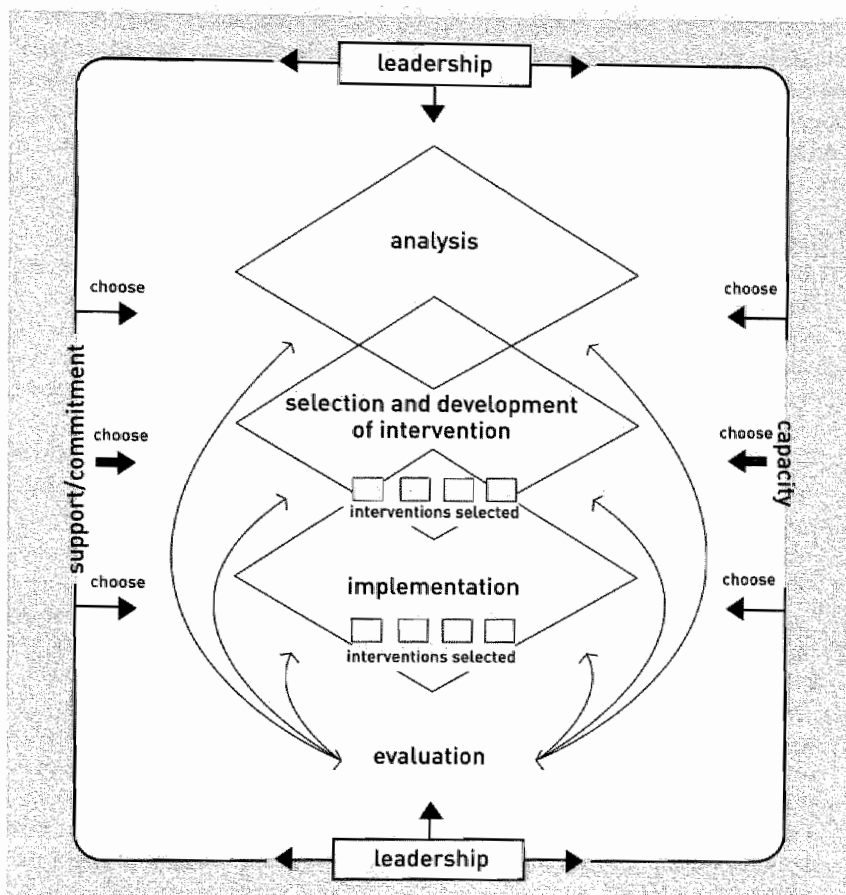


Figure 1: Preffi 2.0 model

clusters on the left-hand side relate mostly to the systematic development of interventions, while those on the right relate particularly to aspects of implementation. The basic structure is the same as that of Preffi 1.0, although certain clusters have been redefined with respect to one another. For instance, the two clusters entitled 'Contextual conditions' and 'Project management' in Preffi 1.0 were replaced by one new cluster, called 'Contextual conditions and feasibility', which, as discussed above, influences all other clusters. Another example is that of the clusters 'Effective techniques' and 'Pretest', which were separate clusters in Preffi 1.0 and were included in the cluster entitled 'Intervention development' in the new version.

At criterion level, Preffi 2.0 was thoroughly revised compared to the first version. The nature and designation of a number of criteria were changed, criteria and clusters were made more rational and theoretically-based, and the number of criteria was reduced from 49 to 39. These changes reflect the new scientific and practical developments and insights that have arisen during the past six years. The main changes are described below.

1. The context of projects was given a more prominent position in the new version, as it is a major factor determining effectiveness. The idea of context was given a broader definition and was specified in greater detail than in Preffi 1.0, which concentrated mostly on capacity. New elements include the level of support available and the leadership by the project manager. These reflect the influence of community projects (in terms of the need to create local support) (Minkler and Wallerstein, 1996; Bracht, 1999; Laverack and Labonte, 2000) and the greater importance currently attached to leadership issues (Groote, 1997; Bekkering, 2001; Saan et al., 2004). In addition, the other criteria now also refer more frequently to contextual conditions and feasibility (for instance criterion 5.4 on the feasibility of objectives).
2. The new version places greater emphasis on the relations between criteria and clusters. In concrete terms, Preffi 2.0 assesses the use of a theoretical model (criterion 3.1) reflecting the relations between determinants and problems (Contento et al., 1995; Godin and Kok, 1996; Glanz et al., 1997; Brug et al., 2000; Bartholomew et al., 2001), the links between objectives and analysis (5.1), the match between intervention methods and objectives or target groups (6.1.a). It also places greater emphasis on the relations between multiple interventions, in that it assesses the coherence of multiple interventions (6.6).
3. Participation and involvement by the target group has in recent years increasingly been emphasised as a major aspect of effectiveness. This was incorporated in Preffi 2.0 by criteria that ask about support/commitment (1.1), problem perception by stakeholders (2.3), motivation and opportunities of the target group (4.2), participation by the target group (6.3.a) and fitting interventions to the target group's culture (6.3.b) (Pretty, 1995; Minkler and Wallerstein, 1996; Bracht, 1999; Huff and Kline, 1999; Laverack and Labonte, 2000; Rifkin et al., 2000).
4. Implementation was given a more prominent role in Preffi 2.0, and was further specified. Whereas Preffi 1.0 concentrated on methodological recom-

mendations (drawing up a plan, training the implementers, etc.), Preffi 2.0 distinguishes between dissemination (and ways of promoting this) and implementation. The new version pays much more attention to the process of dissemination and aspects that are relevant to this process, such as involving the intended implementers at an early stage (6.5) and implementation strategies (7.1.a). It also explicitly mentions implementation interventions (7.1.b) (Paulussen, 1994; Rogers, 1995; Oldenburg et al., 1997; Durlak, 1998; Splunteren and Borselaar, 1998; Splunteren, 1999; Grol and Wensink, 2001).

5. The new version further specifies aspects of progress monitoring and of evaluation. Several criteria ask whether choices have been made (2.3, 3.4, 5.2) to allow evaluation. Pretesting (6.7) and monitoring the progress of the intervention and its implementation (7.2) provide feedback about the progress of a project. In the evaluation cluster, effect evaluation was further specified (8.3), and criteria were added about the principles of evaluation (8.1) and giving feedback on the basis of the evaluation (8.4) (Nutbeam, 1998; Nutbeam, 1999; Swanborn, 1999; Dijker et al., 2000; Francisco et al., 2000; Migchelbrink, 2000; Butterfoss et al., 2001; Koelen et al., 2001).
6. A final new element involves the criteria on 'duration, intensity and timing' (6.2). Since contextual conditions frequently do not allow sustained and labour-intensive interventions, it is important to assess whether an intervention has enough 'substance' to be able to produce changes. There is considerable evidence (Dusenbury and Falco, 1995; Gageldonk and Cuijpers, 1998; Llopis, 2002) that interventions require a minimum exposure to be effective. In some cases, however, short-term interventions may nevertheless be effective, provided they are applied at the right moment, for instance during a 'sensitive period' in someone's life, or during an incident or crisis (Pieterse, 1999; Llopis, 2002).

4.3.3 The scoring method

Each effect predictor (criterion) in Preffi has been operationalised in the form of one or more specific yes/no-type questions. The answers to these questions allow users to rate the degree to which an intervention meets a particular criterion as 'weak', 'moderate' or 'strong' (for scoring form see the foldout flap at the back of the book).

This operationalisation aims to provide Preffi users with an instrument that allows them to assess a programme as objectively as possible. Nevertheless, the

nature of a particular criterion or of the questions operationalising the criterion may not always allow an objective assessment. A rough distinction can be made into three types of criteria. The first type is that of criteria for which the questions can be unequivocally answered, such as 'Is it known to what extent the target group does indeed perceive the problem as a problem?' (2.3), and 'Do the objectives specify the factors to be changed?' (5.2). The second category is that of criteria for which the questions cannot be so straightforwardly answered because the aspect they refer to cannot be so clearly assessed, such as the question about the expertise of the project manager (1.3.a) or the question whether the target group perceives the intervention as compatible with their culture (6.3.b). In such cases, users are advised to seek peer consensus on the answer. Finally, there are a number of criteria which basically require an expert opinion. An example is the operationalisation of 'theoretical model' (3.1), which includes not only the question whether a theoretical model is being used but also asks whether it has been made plausible that the model chosen is suitable for application in a particular situation.

A number of Preffi criteria offer the opportunity to choose the option 'not assessable'; these are particularly those criteria that are difficult to assess by third parties on the basis of documentary evidence provided, such as the expertise and characteristics of the project manager (1.3.a). In the other criteria, a lack of information is assessed as 'weak'. The scoring form allows each cluster to be given an overall mark between 1 and 10, which is to be composed from the individual criterion marks within the cluster. This can ultimately result in an overall mark for the project as a whole, based on the assessments for the individual clusters. The overall mark need not be a simple average, as different weights can be allocated to certain criteria or clusters. The back of the scoring form provides room to enter the cluster scores in graphic format, and to list a limited number of points to be improved and concrete actions required to achieve this improvement.

4.4 Results of the pilot study among practitioners

The pilot study among 35 practitioners (see step 7), which can be regarded as a pre-test for Preffi 2.0, yielded a generally favourable assessment of the draft version and its various components. Of the 28 respondents who had worked with Preffi 1.0 before, 25 regarded Preffi 2.0 as an improvement, particularly because of its improved scientific basis and the operationalisation of the criteria (Meurs, 2002). The instrument was given an average overall mark of 7.7 out of 10, and was evaluated by a large majority as valuable, complete, clear, well-

organised and innovative (see table 1). Most respondents reported Preffi 2.0 to be useful to themselves (83%) and to colleagues (89%), both for project development (79%) and for project evaluation (85%). At the same time, table 1 shows that Preffi 2.0 was regarded as difficult and long, rather than as easy and short. The question asking whether Preffi was pleasant to work with drew a neutral response from 43% of the respondents, while 46% said it was pleasant to work with.

I think Preffi 2.0 is:	5	4	3	2	1	
valuable	54	37	6	3	-	redundant
complete	48	40	9	3	-	incomplete
clear	20	57	14	9	-	unclear
well-organised	20	54	14	9	3	poorly organized
innovative	17	57	23	3	-	out of date
nice to work with	17	29	43	11	-	annoying to work with
easy	-	26	31	40	3	difficult
short	-	-	57	40	3	long
compact	3	-	65	29	3	to extensive

Table 1: opinions of 35 practitioners about the concept version of Preffi 2.0

The participants took an average of 133 minutes to assess the first project with Preffi, against 85 minutes for the second one. The time that users need to invest thus becomes shorter as they become more experienced in its use. The open answers to the questionnaire and the interviews yielded many suggestions for improvements to the way Preffi is used in practice. These have led to a number of adjustments in the instrument's content and new ideas about ways of using Preffi 2.0. The next section discusses these aspects of its use.

4.5 Instructions for using Preffi

4.5.1 Applying Preffi to one's own or other people's projects

The pilot study showed that the practitioners regarded Preffi mainly as an instrument to assess and improve their own projects, on the basis of project documentation. If information on specific aspects required for the assessment is

missing, the practitioners can presumably supplement this with their own knowledge of the projects. Several respondents recommended applying Preffi jointly with colleagues or superiors, to ensure an objective and critical assessment.

Preffi can also be used to assess other people's projects, in contexts like consultation, peer review, supervision or intervention. In such cases, the evidence from documentation will need to be supplemented by a discussion with the project manager and/or others involved in the project. This is even essential when it comes to assessing aspects like support, leadership by the project manager and fitting the project to the culture of the target group, as these aspects can hardly be assessed on the basis of a project description.

4.5.2 Applying Preffi in various project stages and to various types of project

The Preffi development team is of the opinion that Preffi 2.0 can be applied at various moments in the life of a project. At the start of a project, Preffi can be used as a checklist of major aspects that require further specification. While the project is running, the instrument can help users identify strong and weak points and propose improvements. After the project has ended, Preffi can suggest explanations for the findings and identify points where improvements could be made.

Although all clusters included in Preffi should be applied, the order and intensity with which each is applied can be adapted to the type of project. If the instrument is applied to a development project, the emphasis will be on the early stages, that is, the problem analysis and the choices to be made about objectives, target groups and interventions. If the project to be assessed is an implementation project, such as a project for local implementation of an intervention developed by a national agency, then the Preffi assessment may well start with the third stage, that of implementation. Applying Preffi to a community project may start with the criteria on creating support (1.1) and achieving agreement on the problems to be addressed in a project.

The interviews in our pilot study among users confirmed that the precise way in which users actually go through the various project stages depends very much on the type of project and the setting in which it takes place. People frequently go back to earlier stages or anticipate later stages. Sometimes attention is focused temporarily on one of the contextual conditions, before the process of developing or implementing the project continues.

4.5.3 Using Preffi at specific or general level

The pilot study showed that it is important to decide beforehand at what level Preffi is to be applied. Many projects involve multiple interventions, which can be assessed at various levels of specificity. Whether the instrument is to be applied in great detail depends on one's aim: if the aim is to identify specific points for improvement, Preffi should be applied to each individual intervention. If the aim is to get an overall impression, it should be applied at the general level of the project as a whole.

4.5.4 Dealing with missing or unclear information

The study showed that users were often unsure how to answer certain questions in the operationalisation if they lacked the relevant project information, or if the information was unclear or difficult to assess. In many of these cases, users tended to use the 'not assessable' option, which was available for all criteria in the draft version. In the definitive version, we therefore limited the number of criteria for which this option was available to those criteria that are difficult to assess by third parties on the basis of documentation alone, such as the expertise and personal characteristics of the project manager (1.3.a). For the other criteria, missing information should lead to a score of 'weak'. The draft version also included a scoring category of 'not applicable'. Since the study showed that this was rarely used, it was omitted from the definitive version.

4.5.5 Function of Preffi

The present format of Preffi, and the opportunity it offers to assess a project in the form of an overall mark, seem to invite users to apply the instrument for assessment and screening, that is, to choose between projects. Our study showed that opinions about this were sharply divided among practitioners. The participants were unclear about whether Preffi was intended particularly for quality assessment or for quality improvement. The Preffi development team regards the instrument primarily as intended for quality improvement, which means that it should be used particularly for diagnostic purposes. If quality is to be assessed in such a context, the assessment should aim to suggest improvements.

4.6 Conclusions

The systematic process of developing Preffi 2.0 has yielded a quality assurance instrument that is appreciated by both its direct target group, i.e. the health promotion specialists, and by researchers. The instrument combines a sound scientific basis with practical usefulness. This result would not have been possible without the input provided by the researchers and practitioners themselves. In addition, the use of a draft version in a pilot study among potential users generated many new ideas for improvements to its content and appearance and for possible and desirable ways to use it.

Notwithstanding their general appreciation of Preffi 2.0, a substantial part of the target group appears to regard the instrument as somewhat too long and difficult. In addition, the participants in our pilot study encountered a number of problems, particularly in dealing with unclear or missing information in project descriptions. The Preffi development team has tried to remedy or prevent these problems in the definitive version, mostly by clarifying the instructions in the user manual. Moreover, Preffi will in practice be used particularly to assess people's own projects, which means that all the necessary information will be available, whereas the participants in our pilot study were asked to assess other people's projects. Nevertheless, it is clear that using Preffi will require a thorough training.

The operationalisation and norms developed for Preffi 2.0 may suggest that the instrument's focus is more on assessment than on improvement; this is emphatically not what its developers envisaged or wanted. The main idea behind the operationalisations and norms is to allow professionals to analyse their own or their colleagues' projects in a more specific and normative fashion, in the hope that this will enable them to identify more specific points for improvement.

Notes

- a. Dr. Stephan van den Broucke, Flemish Institute for Health Promotion; Prof. dr. Johannes Brug, Erasmus MC; Dr. Willy de Haes, GGD Rotterdam, municipal health services; Dr. Maria Koelen, Wageningen University; Dr. Theo Paulussen, TNO Prevention & Health.
- b. The 53 practitioners were employed by national agencies, municipal health services (GGDs), mental health and addiction care institutes and home care agencies.

Chapter 5 Usefulness and reliability of a draft version of Preffi 2.0

5.1 Introduction

Since the early 1990s, the quality of professional health promotion and disease prevention activities in the Netherlands has been the subject of much research and debate. One major objective of these efforts is to stimulate the use of programmes with the greatest possible effectiveness (Bosma and Hosman, 1990; Veen et al., 1994; Kok and Borne, 1995; Brug et al., 2000; Bartholomew et al., 2001).

The Netherlands Institute for Health Promotion and Disease prevention (NIGZ) started the Preffi programme in the mid-1990s as a long-term process to increase the effectiveness of health promotion interventions through systematic and critical reflection on programmes and projects (Molleman et al., 1995). The core of the Preffi programme is the Preffi instrument itself, which is based on a systematic approach to health promotion projects and programmes. Preffi consists of a set of guidelines with corresponding assessment criteria which are relevant to the effectiveness of health promotion projects. The criteria incorporate research findings on effect predictors as well as practical expertise on effectiveness. Preffi allows users to assess health promotion projects by means of a series of criteria, as well as to suggest options for improvement, indicate their priority status and implement them (Hommels et al., 2000; NIGZ, 2003).

Systematic efforts to promote the implementation of Preffi itself have been made since 1997. As a result, Preffi has acquired a prominent position in Dutch health promotion activities. At the end of 1999, a survey found that Preffi was considered an important instrument, that 96% of Dutch health promotion specialists were aware of its existence, while 73% were familiar with it and about 40% were applying it occasionally or consistently (Molleman et al., 2004).

A 1999 study of the use of Preffi 1.0 by practitioners had shown that Dutch practitioners did not always use a sufficiently systematic approach to various aspects of health promotion (Molleman, 1999; Peters et al., 2000). Although Preffi helped to improve this, it was found that the criteria it included did not offer enough indications to allow users to decide whether they were using a

particular criterion correctly. Users reported that they would like to see this improved in a new version of Preffi, with all criteria operationalised and norms added to each criterion. In addition, new research findings since 1995 also necessitated an updated version.

A systematic project to develop the second version of Preffi was therefore started in 2000. The project was a joint effort with the Prevention Research Centre of Nijmegen University, a scientific advisory board (WAR) including five researchers, and a practitioners' advisory board (PAR) consisting of 45 practitioners. The latter was set up to ensure sufficient input from the users' perspective.

After a thorough process of design and consultation with WAR and PAR, a draft version of Preffi 2.0 was produced in 2002. The reliability and usefulness of this draft instrument were then tested in a study, the results of which are presented in this chapter.

The study tried to answer the following questions:

- Does the draft version of Preffi 2.0 differentiate sufficiently between projects and between criteria within projects, in terms of the presence of effect-promoting conditions in a project or programme?
- Is the draft version of Preffi 2.0 a reliable instrument, in the sense of allowing different practitioners to arrive at the same assessment of health promotion and disease prevention projects?
- Do users consider Preffi 2.0 to be a useful instrument, and what problems do they encounter in applying it?
- What can be done to further improve the reliability and usefulness of the draft version of Preffi 2.0?

5.2 Methods

Thirty-five health promotion specialists (practitioners) were each asked to assess two (from a total of four) project descriptions using the draft version of Preffi 2.0. In addition to completing the Preffi scoring form, they were invited to provide specific comments on the operationalisations of the Preffi criteria. After they had completed the two assessments, they were invited to fill in a questionnaire. In addition, ten of the assessors were interviewed on the basis of their assessments and questionnaire answers, to enable us to gain a more thorough understanding of the use and usefulness of Preffi 2.0.

5.2.1 Study population

We invited 59 practitioners, most of whom had considerable experience, to participate in the study. The practitioners were invited if they were familiar with Preffi 1.0 and had a certain amount of experience with its use. They were deemed to be able to compare the draft version of Preffi 2.0 with Preffi 1.0, and to use the draft version without requiring specific training. In addition to the 45 practitioners who had been actively involved in the development of Preffi 2.0 as members of the PAR advisory board, these included 14 practitioners who were only familiar with Preffi 1.0. No attempt was made to assemble a representative sample. Instead, we tried to compose a group reflecting a wide variety in terms of the major Dutch health promotion sectors, gender, experience, type of position and educational background. We also invited some of the 'opinion leaders' of each sector, insofar as they were not already involved in the PAR.

The participants were told that their participation would require about 4 to 6 hours of their time. In the end, 42 of them were willing to take part, of whom 28 were PAR members. Ten of the participants were also interviewed. These were selected in such a way as to include various positions (including project managers and directors of health promotion departments) and sectors, and to include both PAR members and non-members.

5.2.2 Measurement instruments

The reliability and usefulness of Preffi were tested using four descriptions of existing projects, which were all nominated for the 'Preffi award' in recent years (Molleman et al., 2000). The Preffi award is the bi-annual award for the best health promotion project or programme in the Netherlands. The subjects of these four projects were healthier diets, accident prevention, depression and the prevention of bullying. Clear project descriptions were available, which needed very little adaptation to allow data to be collected on all 38 Preffi criteria. The first two of the projects derived from municipal health services, the other two from prevention services in mental health.

The project team itself assessed one of the projects by way of pre-test. The results were used to clarify a number of aspects in the project descriptions. Each assessor was invited to assess one project relating to municipal health services and one relating to prevention services in mental health. The projects were divided at random among the assessors, who were presented with them in different orders.

The data for the reliability test were scored on the draft Preffi 2.0 scoring form. Respondents were asked to assign a score of 'strong', 'moderate', 'weak', 'not assessable' or 'not applicable' for each of the 38 Preffi criteria for each project. In addition, they were invited to assign marks between 1 and 10 for each of the 8 clusters of criteria and for the project as a whole. Finally, they were asked to indicate points where the projects could be improved and the actions required to achieve these improvements.

Additional information was obtained from the questionnaires and interviews, where we asked the participants to explain the considerations which had determined their scorings.

Usefulness was examined using a number of instruments. The first was a questionnaire. This asked about the respondents' background variables, such as the type of organisation they worked for, their position, their educational background and their familiarity with Preffi 1.0. It also asked them whether they were familiar with the projects to be assessed, and how much time they had spent assessing them, as well as their opinions on a number of aspects of usefulness (degree of difficulty, clarity, completeness, innovativeness, etc), for Preffi as a whole and for its constituent parts. They were further asked whether they intended to start using Preffi 2.0, how they appreciated the new version and whether they had any suggestions for improving the instrument.

The respondents were also invited to use the margin of the 'Operationalisation and Norms' document to write specific comments and suggestions for improving the operationalisation descriptions.

The interviews were used to discuss the respondents' experiences with the use of the draft version of Preffi 2.0, as well as to allow them to clarify any unusual scores and to ask them how they might use Preffi 2.0 in practice and how it could be further improved.

5.2.3 Data analysis

To answer our first research question (i.e., whether the draft version of Preffi 2.0 adequately differentiates between and within projects in terms of the presence of effect-promoting conditions in a project or programme) we compared the scores assigned to the projects. We examined which items were hard to assess and whether there were any salient differences in the scores assigned per criterion, cluster or project. We also studied the correspondence between the scores and the suggestions for improvement that were made.

The 'not assessable' option was analysed in some more detail. This option was recoded afterwards, using a conservative approach: if a project description provided no information on a particular Preffi criterion, such as the theoretical framework used for the project, we assumed that this aspect had been insufficiently incorporated in the design and implementation of the project. In such cases, the user manual recommends a score of 'weak' rather than 'not assessable'. There are a number of exceptions to this, relating mostly to those Preffi criteria that ask the assessor to judge certain management aspects of the project. Such aspects include the characteristics of the project manager or the adequacy of the budget. If the project description failed to provide information about these aspects, we recommended a neutral score of 'moderate'. This recoding was particularly required to allow us to do analyses based on generalisability theory.

The scores assigned to the criteria were compared with those at cluster level to check whether the respondents had given greater weight to some criteria than to others in arriving at a cluster score. We also checked whether they had applied weighted scores at cluster level to arrive at an overall project score.

The reliability of the assessments (our second research question) was analysed using generalisability theory (Shavelson and Webb, 1991; Eggen and Sanders, 1993). This theory, an extension of the classical test theory, addresses the question how accurately the criterion scores can be generalised across the factors contributing to the scores. It checks the extent to which differences in measurements (i.e., the scores assigned per criterion, per cluster or to the entire project) can be attributed to true score components (in this case the characteristics of the project scores) or to potential error sources, in this case the assessors and the measurement instrument at criterion and cluster level. These components of variance were estimated using the varcomp procedure in SPSS.

These were in turn used to calculate the G (generalisability) coefficient and the SEM (Standard Error of Measurement). The G coefficient is a measure of reliability, indicating the ratio between the true variance and the total variance. The SEM allows a confidence interval around a score to be calculated, making it a measure of the accuracy of the measurement procedure.

A reliability study using generalisability theory consists of two steps. The first step, the Generalisability study (G study), was described above. This is followed by a Decision study (D study) to determine the number of assessors needed to ensure reliable conclusions.

The usual minimum reliability threshold for the G coefficient is 0.70 (Nunnally, 1967). There is no generally accepted maximum value for SEM. A reasonable

requirement would seem to be that the confidence interval of the score remains within the rounding zone for the score category, that is, half a point above or below the category. In other words, the confidence interval around a score of 2 does not include scores lower than 1.5 or higher than 2.5. To achieve this level of precision, SEM needs to be smaller than 0.26. (See also appendix 2 for the equations used to calculate G and SEM).

Since each assessor assessed two of the four projects, the assessors were nested within the projects. The assessors and projects were random facets, since the total number of projects and assessors is theoretically infinite. An additional source of variance in the analyses at cluster level is that of the criteria, which were regarded as fixed facets, as we did not need to generalise to a universe of criteria.

The comments on operationalisations, the answers to the open questions and the contents of the interviews were collected and summarised (Meurs, 2002). They provided the building blocks for an answer to our fourth research question, that is, what can be done to further improve the reliability and usefulness of the draft version. They supplied many concrete suggestions for changes to the wording of the operationalisations and norms, as well as input for the user manual that was to be produced. They also provided some clues that helped to explain the findings.

5.3 Results

5.3.1 Response

Of the 42 practitioners we invited to participate, 35 completed and returned the two scoring forms and the questionnaire; a response rate of 83%. Those who did not respond indicated lack of time as a reason for not completing the forms. Non-response was restricted to those persons who were PAR members, especially those working in municipal health service. No significant differences in relevant aspects of the assessments were found between members and non-members of the PAR advisory board, between men and women, between those experienced in using Preffi and those less experienced or between respondents working in different sectors.

The respondents had an average of 13 years of experience in health promotion and were usually project managers or directors of health promotion depart-

ments in organisations. Eighty percent had worked with Preffi 1.0. This means that the study population was more experienced and had higher positions than the average Dutch health promotion practitioner.

5.3.2 Project assessments

The respondents' average scores for individual criteria provided a useful overview of each project's strong points and those aspects which leave room for improvement. Since a detailed discussion of the data per project is beyond the scope of this chapter, we limit ourselves to some general comments about the project scores.

The 'not assessable' option was used 394 times (14.8%). The comments provided by the assessors when using this option showed that they felt the project descriptions gave too little information on the particular criterion. The use of this option was not evenly distributed across the clusters and criteria. The average percentage of 'not assessable' scores in the clusters that include criteria relating to a project's development aspects (Preffi clusters 2-6) was 9%. The criteria within the intervention development cluster for which this option was used most frequently include particularly those newly developed for Preffi 2.0, such as intensity, timing, 'fitting to culture' and 'feasibility in existing practice'. The average percentage of 'not assessable' scores in the clusters including criteria relating to the implementation (contextual conditions, implementation and evaluation) was 24.2%. The highest average percentage within these clusters was found for contextual conditions (31%), while those for implementation and evaluation were about 22%. Exceptionally high percentages were those for criterion 1.3a (expertise and characteristics of the project manager) with 60%, 8.4 (feedback to stakeholders) with 40% and 7.1c (appropriateness of the supplier for intermediaries) with 34%. These are all aspects that were hard to judge on the basis of the project descriptions.

Of the 38 Preffi criteria assessed, 18 were already included in Preffi 1.0, while the remaining 20 were new. The results show that the 'not assessable' option was used significantly more often for the new criteria than for those already included in Preffi 1.0 ($t=-2.551$ $df=32.6$ $p=.016$).

The option 'not applicable' was used 25 times (0.9%). In 7 of these cases, this related to criterion 6.2b, 'timing of the intervention', in 4 cases to the pre-test (6.6) and in 3 cases to the criterion on feedback about the evaluation to stakeholders (8.4). The remaining cases all related to criteria for which this option was used only once.

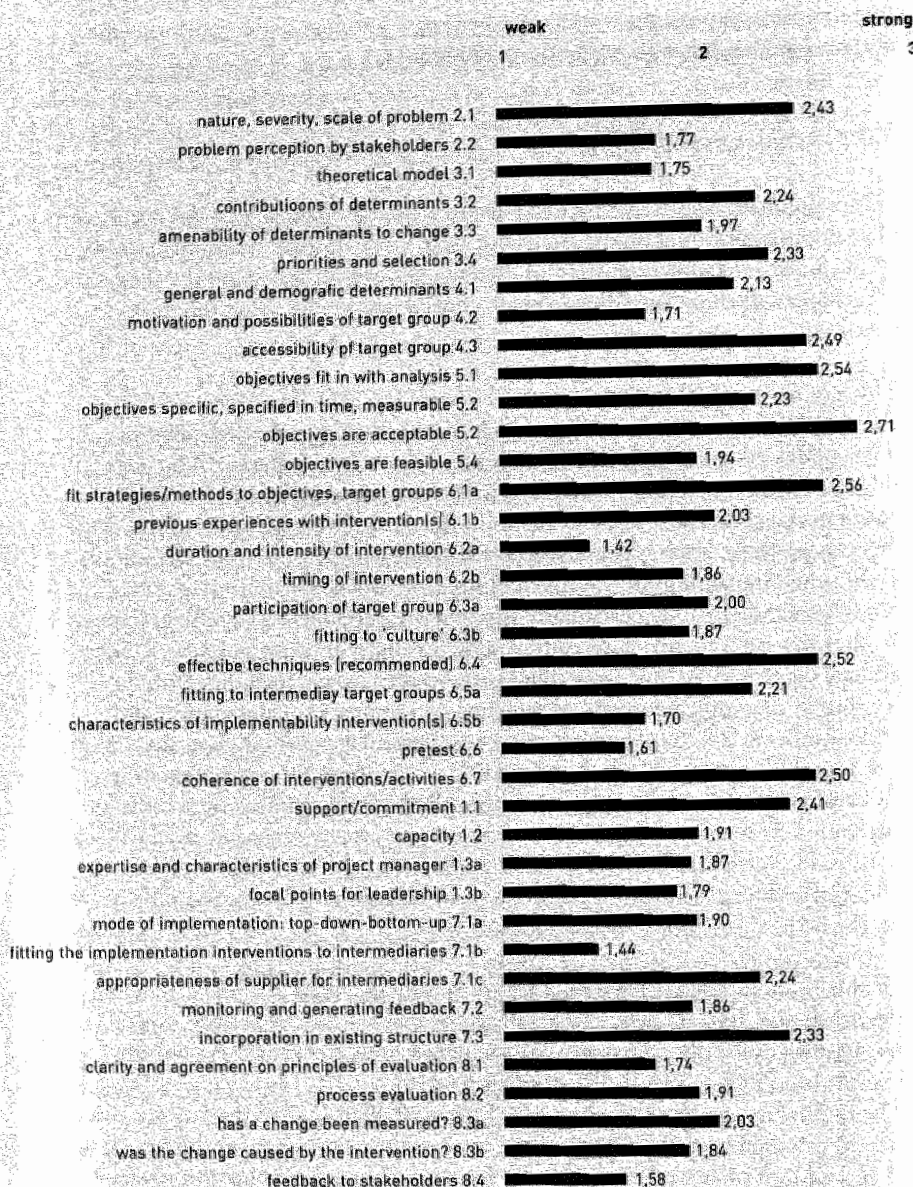


Figure 1: the average criteria scores for all four projects together

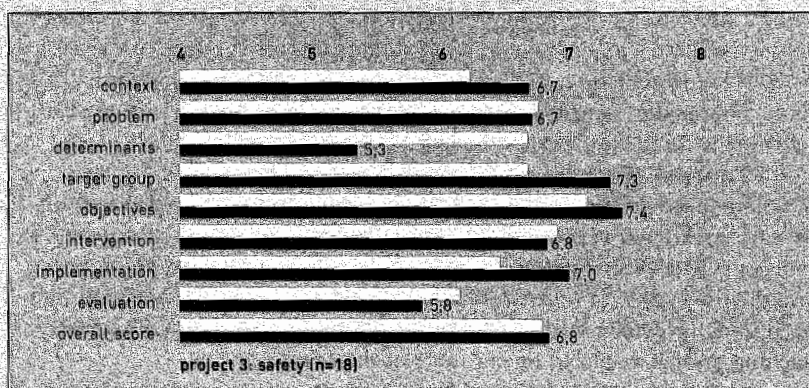
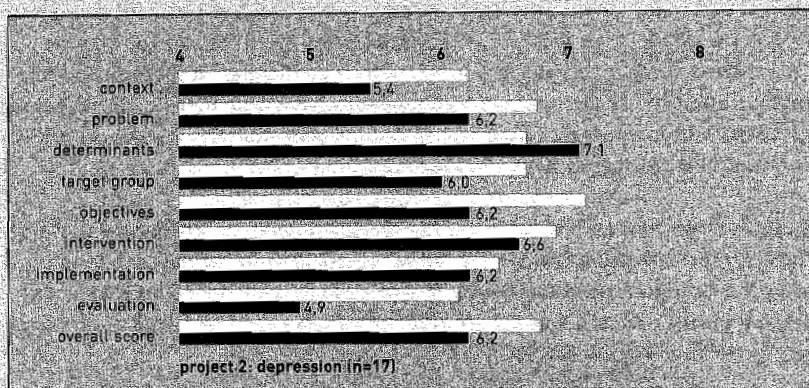
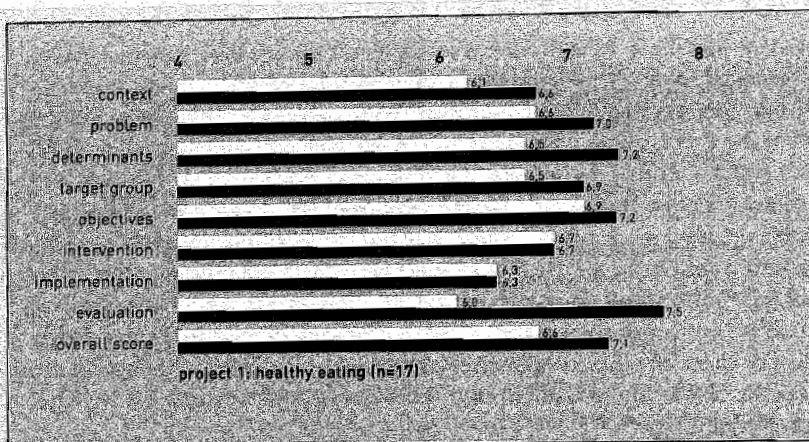
Figure 1 shows the average criteria scores for all four projects together, using the recoded scores. The figure shows what a project assessment can look like. All four projects were of a high quality, or they would not have been nominated for the Preffi award. The strengths (i.e., criteria scoring higher than 2.25 on a scale of 1 – 3) that these projects had in common were their thorough and systematic design (high scores for the analysis of the nature and scale of the problem; clear priorities, with objectives and interventions fitting in with these priorities; coherence between interventions), their use of effective techniques and their suitability for the relevant context (in terms of acceptable objectives, accessible target groups, support and incorporation in an existing structure). Most of these 10 criteria were already included in Preffi 1.0, although the context criteria were more explicitly defined in the second version, and were thus in fact new.

The 8 criteria that yielded the lowest average scores (below 1.75) for these four projects included 7 that were newly introduced in Preffi 2.0. These include items such as using a theoretical model, motivation of the target group, duration and intensity of the intervention, an explicit implementation strategy and a clear evaluation strategy. These reflect new insights and are apparently not yet commonly included in project descriptions.

Figure 2 shows the cluster scores and overall project scores for each of the four projects, plotted against the average scores across the four projects. The figure shows the strengths and weaknesses of each project, which implies that assessment by means of Preffi was able to differentiate between and within projects, even for relatively good projects.

The assessors also made suggestions for improvements to the projects. Suggestions for project 1 related to implementation, those for project 2 to evaluation, those for project 3 to analysis and evaluation and those for project 4 to evaluation and implementation. These suggestions corresponded well with the cluster scores. The assessments of these four projects also largely agree with the comments made on the projects by the jury of the Preffi award.

There were no significant differences in the absolute scoring levels between assessors working in the various types of organisation. The assessments reflect a tendency whereby the assessors passed more critical judgements on projects from their own type of organisation, but the differences were not significant. At group level, no evidence was found of the option to give weighted scores by attaching greater weight to one or more of the criteria.



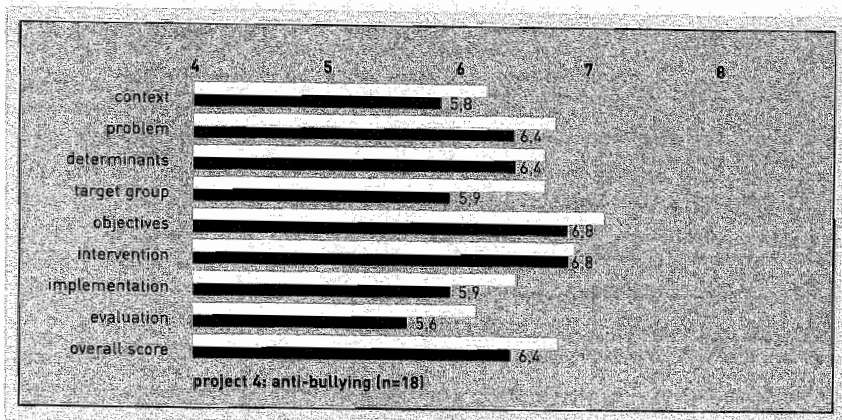


Figure 2: scores per cluster and overall project score for each of the projects (NB. Upper (light) bar always represent the average for all four projects together; numerical values of averages scores are only indicated in the 'healthy eating' panel (n=70))

5.3.3 Reliability

Table 1 lists the variance estimates for each criterion. The table shows the percentage of the total variance contributed by each source of variance (the project and the assessor nested within the projects). The variance component for the project is an estimate of the true score variance, which should be as high as possible. Since the variance component relating to the assessor nested within projects (*ass:proj*) corresponds to the component relating to the interaction between project and assessor, all variance not attributed to the project has been included as *ass:proj*.

At cluster level and the level of the project as a whole, there is also the criterion facet. Since this is regarded as a fixed facet, however, this does not add any variance components to the table.

The table shows that the variance between the projects was small or absent for many of the criteria. By definition, this means a low G coefficient, since reliability is defined as the ratio between the true score variance and the total variance. In those cases where there was some variance between the projects, the minimum G value of .70 was indeed achieved. Most of the variance resulted from the interaction between project and assessor. An analysis at cluster level shows that 5 of the 8 clusters also yielded too little variance to obtain an acceptable G value. The same is therefore true for the total of all criteria, which showed a variance of only 5.5% between the projects.

		G abs	SEM	% of total variance		SEM < .26	SEM
	number of assessors	17	17	project	ass-proj	n=	for n=4
2	Problem analysis	.07	.13	0.4%	99.6%		27
2.1	Nature, severity, scale of problem	.00	.15	0.0%	100.0%	6	
2.2	Problem perception by stakeholders	.27	.19	2.1%	97.9%	9	
3	Determinants	.87	.11	28.5%	71.5%		25
3.1	Theoretical model	.84	.17	23.4%	76.6%	8	
3.2	Contributions of determinants	.47	.12	4.9%	95.1%	4	
3.3	Amenability of determinants to change	.86	.19	25.8%	74.2%	9	
3.4	Priorities and selection	.83	.17	22.7%	77.3%	8	
4	Target group	.76	.12	15.8%	84.2%		25
4.1	General and demographic characteristics	.00	.19	0.0%	100.0%	10	
4.2	Motivation and possibilities of target group	.89	.17	31.8%	68.2%	7	
4.3	Accessibility of target group	.23	.17	1.7%	98.3%	8	
5	Objectives	.65	.11	10.0%	90.0%		22
5.1	Objectives fit in with analysis	.00	.16	0.0%	100.0%	7	
5.2	Objectives: specific, specified in time, measurable	.93	.16	42.8%	57.2%	7	
5.3	Objectives are acceptable	.31	.14	2.5%	97.5%	5	
5.4	Objectives are feasible	.00	.21	0.0%	100.0%	12	
6	Intervention development	.55	.09	6.6%	93.4%		.25 [n=2]
6.1a	Fit strategies/methods to target, target groups	.58	.13	7.6%	92.4%	5	
6.1b	Previous experiences with intervention(s)	.67	.20	10.5%	89.5%	10	
6.2a	Duration and intensity of intervention	.81	.17	20.0%	80.0%	8	
6.2b	Timing of intervention	.88	.18	30.3%	69.7%	9	
6.3a	Participation of target group	.94	.17	46.2%	53.8%	8	
6.3b	Fitting to 'culture'	.14	.21	0.9%	99.1%	11	
6.4	Effective techniques [recommended]	.03	.14	0.2%	99.8%	5	
6.5a	Fitting to intermediary target groups	.67	.20	10.4%	89.6%	10	
6.5b	Characteristics of implementability intervention(s)	.00	.20	0.0%	100.0%	10	
6.6	Pretest	.49	.19	5.4%	94.6%	10	
6.7	Coherence of interventions/activities	.22	.15	1.6%	98.4%	6	
1	Contextual conditions and feasibility	.43	.12	4.2%	95.8%		25
1.1	Support/commitment	.81	.18	19.7%	80.3%	8	
1.2	Capacity	.32	.20	2.7%	97.3%	11	
1.3a	Expertise and characteristics of project manager	.41	.14	3.9%	96.1%	5	
1.3b	Focal points for leadership	.37	.15	3.4%	96.6%	6	
7	Implementation	.49	.13	5.3%	94.7%		26
7.1a	Mode of implementation: top-down - bottom-up	.89	.18	32.4%	67.6%	8	
7.1b	Fitting the implementation interventions to intermediaries	.00	.18	0.0%	100.0%	9	
7.1c	Appropriateness of supplier for intermediaries	.06	.16	0.4%	99.6%	7	
7.2	Monitoring and generating feedback	.60	.21	8.2%	91.8%	12	
7.3	Incorporation in existing structure	.43	.19	4.3%	95.7%	9	
8	Evaluation	.86	.13	26.6%	73.4%		26
8.1	Clarity and agreement on principles of evaluation	.89	.18	31.2%	68.8%	9	
8.2	Process evaluation	.77	.19	16.0%	84.0%	9	
8.3a	Has a change been measured?	.92	.15	41.3%	58.7%	6	
8.3b	Was the change caused by the intervention?	.73	.18	13.5%	86.5%	8	
8.4	Feedback to stakeholders	.61	.20	8.4%	91.6%	10	
Total project (values for all criteria together)		.50	.08	5.5%	94.5%		24 [n=2]

Table 1: G and SEM values for each individual item, per cluster, and for the assessments as a whole, plus the numbers of assessors needed to achieve an accurate assessment, per item, and the sem score per cluster with 4 assessors

The Preffi instrument was assumed to yield accurate conclusions if the standard error or measurement (SEM) was smaller than .26. Calculation of this value across the 17 assessors for each project showed that the assessments for each of the criteria were sufficiently accurate.

The variance components can be used to calculate how many assessors are required to meet the minimum G and maximum SEM values. Due to the low variance at criteria level, the number of assessors needed to achieve the minimum G value cannot be meaningfully calculated. Table 1 lists the number of assessors required to achieve an acceptable SEM value for each criterion. This number ranged from 4 to 12. Those criteria requiring 7 or fewer assessors can be regarded as relatively strong.

Combining the criterion-level assessments to cluster level shows that at this level, 4 assessors were enough to achieve an SEM equal to or lower than .26. This was true for all clusters except that of problem analysis. This means that 5 assessors would be enough to produce sufficiently accurate assessments at cluster level. At the level of the projects as a whole, assessment by 2 assessors yielded an SEM value of .24.

		G abs	SEM	% of total variance		SEM < .26
number of assessors		17	17	project	ass:proj	for n=
2	problem analysis	.06	.34	0,4%	99,6%	29
3	determinants	.81	.37	19,8%	80,2%	35
4	target group	.77	.32	16,4%	83,6%	26
5	objectives	.63	.36	9,0%	91,0%	32
6	intervention development	.00	.32	0,0%	100,0%	26
1	contextual conditions	.66	.36	10,3%	89,7%	34
7	implementation	.31	.40	2,5%	97,5%	40
8	evaluation	.86	.41	26,1%	73,9%	43
	overall project score	.20	.31	1,5%	98,5%	26
values for all clusters together *		.54	.27	6,4%	93,6%	18
* clusters scores are considered fixed facets						

Table 2: reliability (G) and accuracy (SEM) values of assessment scores per cluster and for the project as a whole plus the numbers of assessors needed to achieve an accurate assessment, per cluster and for the project as a whole

Table 2 lists the G and SEM values for the *overall marks assigned per cluster* and for the *project as a whole*. This table once again shows the low variance between the projects. Where the variance was sufficient, G values of $> .70$ were found. None of the overall marks had an SEM smaller than .26. The use of a 10-point scale for these cluster and project marks means that accuracy is likely to be smaller than for the 3-point scales used to assess the individual criteria. This means that larger numbers of assessors are needed: between 25 and 40 assessors would be required to produce adequate accuracy at the level of the overall cluster marks.

In their comments in the margins of the operationalisation document, as well as in the questionnaires and interviews, the respondents reported a number of problems they had encountered while assessing the projects. Solving these problems could contribute to a greater accuracy in Preffi assessments. The problems mentioned included the following:

- a. The project descriptions often fail to provide information about aspects that the operationalisations refer to.
- b. Many respondents were unsure about the distinction between the scoring options 'weak' and 'not assessable'.
- c. The operationalisations occasionally use words like 'adequate' and 'sufficient', which leave room for interpretation by individual assessors.
- d. Some respondents reported that the operationalisations had made them more critical in their assessments. Others had applied the operationalisations in a less critical manner as the operationalisations sometimes led to scores that they felt were too low for the projects concerned. Some respondents reported that being familiar with a similar project made them more lenient in their assessment, since it meant that they knew how difficult these things can be to set up, while others claimed that such familiarity made their assessment stricter, as it meant they were better aware of the nature of the problems, and were therefore able to apply the operationalisations more stringently. However, our further analyses of the degree of familiarity with the projects, the degree to which respondents felt influenced by this and their general level of job experience did not reveal any significant relations with the marks assigned.
- e. Many respondents commented on the method used to score multiple interventions within one project. The present Preffi scoring form only allows users to assign one score to each set of interventions, even though there may be large differences between the scores for individual interventions. The score assigned thus becomes a general average, which is open to various interpretations.
- f. Respondents found it difficult to assign an overall mark per cluster or to the project as a whole when much information was lacking from the project descriptions.

5.3.4 Usefulness

Of the 35 respondents, 28 had already worked with Preffi 1.0. Of these, 25 felt that the draft version of Preffi 2.0 represented an improvement, or even a great improvement over Preffi 1.0. The main arguments mentioned included the new operationalisations and norms for the criteria, and the scientific rationale for the criteria provided in the document entitled 'Explanatory Guide'. The respondents also felt that Preffi 2.0 'was better suited to the dynamic nature of practical situations'. The use of Preffi stimulated them to be more critical in their assessment and to screen all aspects of a project, which allowed them to quickly identify points for improvement. When asked to give Preffi 2.0 an overall mark between 1 and 10, the respondents assigned it an average mark of 7.7.

The average time respondents needed to assess the first project was 113 minutes ($SD=46.7$), while the second took 83 minutes ($SD= 36.6$). There were, however, large differences between respondents in this respect: the time required ranged from 34 minutes for the fastest respondent to 245 for the slowest. No significant relation was found between the time required and the amount of experience or the sector the respondents were working in. The respondents regarded the time they had invested as reasonable. Several respondents reported that it had taken them some time to familiarise themselves with the Preffi terminology, but that the second assessment was considerably easier and faster. It is our expectation – and experiences so far seem to bear this out – that an assessor with some experience in using Preffi should be able to assess a project in an hour.

Twenty-six percent of the assessors were familiar with the topic of a project they were asked to assess or had been involved in similar projects on the same topic. The assessors indicated that this had influenced their assessment ($\chi^2 = 27.89$, $df=10$, $p=.002$). Familiarity with a project might have a positive or negative effect on the judgements. There was no significant difference in overall marks for clusters or the project as a whole between those who were familiar with a project and those who were not or barely familiar with it.

The study evaluated all components of the new Preffi version on a number of dimensions. The resulting overall impression of Preffi 2.0 as a whole is that of a valuable, comprehensive, clear and innovative instrument (table 3). Table 3 also shows, however, that Preffi 2.0 was regarded as difficult and long rather than as easy and short. In terms of the unpleasant / pleasant dimension, 43% gave a neutral score, while 46% considered the instrument pleasant to use. The

I think Preffi 2.0 is:	5	4	3	2	1	
valuable	54	37	6	3	-	useless
complete	48	40	9	3	-	incomplete
clear	20	57	14	9	-	unclear
well-organised	20	54	14	9	3	poorly organised
innovative	17	57	23	3	-	obsolete
pleasant to work with	17	29	43	11	-	annoying to work with
easy	-	26	31	40	3	difficult
too short	-	-	57	40	3	too long
too compact	3	-	65	29	3	too extensive

Table 3: opinions of 35 practitioners about the draft version of Preffi 2.0

scoring form and the 'Operationalisation and Norms' document were given evaluations similar to or more positive than the general impression. The 'Explanatory Guide' was regarded as useful or very useful by 82% of the respondents. Evaluative scores on this document in terms of the easy / difficult dimension were neutral. It was regarded as long and comprehensive. About 30% of the respondents had hardly used the Guide, or not at all. Many respondents used the opportunity provided by the 'Operationalisation and Norms' document to enter notes and comments on the projects. The interviews also showed that many felt this to be a useful feature. The notes helped the respondents to motivate the scores they had assigned on the scoring form and to discuss these later.

Respondents gave the least favourable opinion about the 'web diagram' that was originally included to allow users to represent cluster scores graphically. Most respondents regarded this feature as unclear and not easy to work with. Some commented that such a graphic representation would only be useful if the scores could be plotted against a standard, against similar projects or against a new assessment of the same project at a later stage.

In addition to evaluating the various components of the draft version of Preffi 2.0, many respondents added suggestions for improvements. Many of these could be incorporated in a more comprehensive new edition of the User Manual. The respondents also criticised Preffi 2.0 for having become more complex and time-consuming than the first version. The wording of some of the questions in

the operationalisations was reported to be too difficult, and the norms too coercive. Some regarded the new version as too academic, and commented that this will not facilitate its acceptance and implementation.

Respondents found it difficult to assign overall marks to clusters and a project as a whole. Several respondents were afraid such marks would be used out of context and thus reduce the value of the Preffi assessment as a whole.

Some suggestions were made to simplify the new Preffi version. Several respondents recommended a digital version, which would make the instrument more manageable. This would allow users to consult specific relevant information and add specific suggestions for improvement to the scores for particular criteria, which would allow the instrument to be used more rapidly and thoroughly.

Some proposed to create different routes through Preffi for different types of projects. Respondents expressed a need for a quick-scan option to lead users to relevant components of Preffi, without the need to apply the entire instrument. Respondents also expressed a desire for a clear user manual and a practical handbook for Preffi users, which explains the content of the present Explanatory Guide to practitioners.

5.3.5 Applying Preffi

Of the 35 respondents, 30 indicated that they would definitely or probably use the new Preffi version in their work. Nearly all felt that Preffi was a useful instrument for practitioners, and with a few exceptions, they also regarded it as useful for themselves. The exceptions were some respondents who felt they were experienced enough not to need an instrument like Preffi as an additional stimulus to work more effectively. Preffi was furthermore regarded as valuable in setting up projects, providing ideas during the design stage, and for regular evaluation of the progress of a project. Preffi could also be used later to evaluate a project after its completion.

The answers to the open questions, as well as the interviews, frequently yielded the suggestion to apply Preffi together with colleagues. The completed scoring form could provide a useful basis for consultations about a project with the director of the health promotion department and/or colleagues. If the scoring form is completed on the basis of a project description, consultation with the project manager will often be required to obtain and evaluate the information not included in the description. At the same time, some respondents doubted

whether enough time would always be made available in practice for such a procedure. This means that Preffi will be used in a flexible manner, with users selecting those components that are relevant at that particular stage.

The scoring form design used in Preffi suggests that it is intended as a measurement instrument that allows projects to be unambiguously and objectively assessed, that is, as an assessment instrument. The basic question is whether this would indeed be possible, or whether Preffi is rather a quality assurance instrument that suggests improvements. The interviews revealed a certain reluctance to use Preffi as an assessment instrument. Several respondents emphasised that Preffi ought to be regarded as a quality improvement tool, and that people should think carefully before using it as a screening instrument. There is a considerable risk that third parties might see it as an assessment instrument and use it to accept or reject projects. The user manual will have to warn about this explicitly.

Some interviewees expressed doubt as to whether Preffi could be used for community projects. They felt that Preffi could only be used selectively, for certain components of these programmes. In addition, some felt that Preffi placed too much emphasis on behavioural change, and not enough on other forms of health promotion, such as legislation and changing people's environment.

5.4 Discussion and conclusions

The draft version of Preffi 2.0 was designed by means of a careful development process, in close consultation with researchers and users. The present study intended to test the practical application of the instrument. Its findings show that this draft version of Preffi 2.0 allows a useful analysis to be made of the strengths and weaknesses of a project. This means that Preffi achieves its intended objective of identifying weaknesses and proposing suggestions for improvement. In addition, the assessments it produces at cluster level are sufficiently accurate provided 5 assessors are used.

The present study also showed that it is difficult to assess a project purely from a written project description. The study has yielded much valuable information on the use of Preffi and its appreciation by users, as well as suggesting ways to further improve the instrument's usefulness and reliability.

Before discussing this in detail, some limitations of the present study need to be addressed. Our decision to use project descriptions as the object of study led

us to look for projects that were adequately described. This search yielded projects that had been designed on the basis of a process of careful consideration; they were among the best projects nationwide and some had already been partly completed. This meant that our respondents applied Preffi to a selected group of projects. As a consequence, many aspects of these projects produced less variance in the assessments than might have been the case if we had chosen a representative sample of Dutch projects. This lack of variance automatically led to a low G coefficient for the relevant criteria and clusters. The G coefficient for those criteria that showed sufficient variance did exceed the minimum value of .70 that allows an instrument to be regarded as reliable. Before Preffi as a whole can be regarded as a reliable instrument, however, it will have to be tested on a set of projects that is more representative of the heterogeneous total population of projects.

In our view, the decision to use high-quality projects for the present study did not affect the opinions of the assessors about Preffi's usefulness as an assessment instrument for projects. It is possible, though, that the use of Preffi may cause users to become annoyed when assessing poor quality projects, as it brings out the shortcomings more clearly.

Our study design was an uneasy compromise between its twin aims of assessing the instrument's reliability and its usefulness. Whereas the inter-assessor reliability might have been tested more adequately by having fewer respondents assess a larger number of projects with a wider range of quality, the usefulness assessment would have benefited from a larger range of assessors.

The usefulness of the Preffi instrument is a major aspect in deciding its value both as a quality improvement instrument and as an assessment instrument. Although its reliability is also important for both functions, it is particularly essential when it is used to assess projects. It is less important for quality improvement applications, as these concentrate on identifying points for improvement rather than providing absolute assessments for the various criteria.

With hindsight, it might have been better to assess these two aspects, reliability and usefulness, in separate studies. In any case, the results made us aware of a number of problems that were taken into account when designing the further study (see chapter 6).

5.4.1 Improving the instrument

The feedback obtained from the present study has been used to make certain adjustments to the definitive version of Preffi 2.0 (Molleman et al., 2003; Peters et al., 2003). These changes, which relate to the scoring form, the operationalisations and the user manual, are briefly described in table 4. It is our hope that this has eliminated a number of problems that the respondents in the study encountered.

We have considerably limited the number of criteria for which 'not assessable' is an option on the scoring form, in order to force users to make as many clear assessments as possible. Reducing the number of scoring options will raise the instrument's reliability.

The revised User Manual now discusses a number of drawbacks and problems our respondents encountered while using the draft version.

The main item of debate as regards the usefulness of Preffi 2.0 is the nature of the instrument, that is, whether it is a quality assurance instrument intended to improve the quality of preventive interventions, or an assessment instrument that allows projects to be selected for implementation. Practitioners will mostly use Preffi 2.0 to assess their own projects, which means they have all the necessary information available. The Preffi development team therefore emphasise the use of Preffi as a quality assurance instrument to improve the effectiveness of existing health promotion projects.

If Preffi is to be used in assessments for selection purposes, the instrument and its application will need to be further developed, to allow reliable assessments to be achieved with only a few assessors. Nevertheless, the new Preffi version does set a certain standard, which it is desirable to meet.

One of the comments of the respondents was that the present version of Preffi concentrates too much on individual behavioural change, and pays too little explicit attention to the importance of environmental determinants and a comprehensive approach to problems, as would be preferred in a broader interpretation of health promotion. In practical situations, they feel, support by the community is found to be such an essential factor that it is impossible to plan interventions in advance, which is another reason why they thought this version of Preffi was less suitable for a community approach. The development team does not agree with this criticism.

Scoring form

- The scoring option 'not applicable' has been removed; only 12 criteria now offer the option 'not assessable'.
- The 'spider's web' has been replaced by a horizontal graphic representation, with three ranges: weak (1 – 4) and amenable to improvement; moderate (5 – 7) and amenable to improvement depending on circumstances; and good (→7)

Operationalisations and norms

- The margin in the operationalisations part has been retained as users found it useful to enter comments about the scores that could be used in later discussions. The operationalisations now allow each subquestion to be ticked as 'yes' or 'no'.
- Questions not relevant to the norms have been removed, as they caused confusion.
- Numerous suggestions made by the respondents to improve the wording have been incorporated in the final version.
- The original criterion 2.1 has been split up into 2 criteria by creating a separate criterion for distribution; additional questions deleted.
- The wording of cluster 3 now emphasises that the determinants of the problem may also relate to psychological problems.
- Criterion 4.1 now also includes size of the target group and the number of characteristics that need to be clear.
- Criterion 5.1 is a crucial question. The levels of the objectives have been added.
- Criterion 6 now starts with an introduction indicating the options of scoring all interventions separately or give an overall assessment. This remains problematic.
- The wording of the question in 6.2 has been clarified.
- The order of criteria 6.6 and 6.7 has been reversed, so that assessors have to score coherence first and then pretesting.
- Identifying the implementation strategy (top-down or bottom-up) was found to be difficult. The criterion now first asks whether intermediaries can adapt the intervention to their own situation and then asks about the essential elements, that must be retained if they want to adapt the intervention to their own context.
- The various stages are now described in 7.1.b, as these were new to many of the respondents.

User Manual

- The Manual has been expanded from 2 to 13 pages, and has been included as the first part of the assessment package. The manual now discusses in detail all components of the scoring form, the use of the operationalisations and norms, the instructions for users and the relation between these documents and the Explanatory Guide.
- The Manual now suggests special routes through Preffi for specific applications.
- The Manual now instructs assessors to make a clear choice of assessing one intervention or more interventions.
- The Manual now emphasises the use of Preffi as a diagnostic instrument for the improvement of projects. Users are encouraged to discuss their assessments with others.

Table 4: Alterations incorporated in the final version of Preffi 2.0, based on the findings of the reliability and usefulness study

Several components of Preffi do try to pay attention to a broader approach to health promotion, for instance the problem analysis and the analysis of determinants, which explicitly advocate a broad approach to problems. Similarly, the specific attention paid to support and commitment by the target group as a factor in intervention development is intended to allow for activities that are part of community programmes. Preffi thus makes allowance for the specific dynamics of this type of process within health promotion programmes. At the same time, we feel that community programmes also require systematic and methodical assessment to improve the effectiveness of their interventions. As we have so far apparently failed to communicate this aspect of Preffi adequately, the revised Explanatory Guide and User Manual will feature clearer explanations of these aspects. One useful new aspect could be the use of specific routes within Preffi.

5.4.2 Improving reliability

It is our expectation that the above improvements to the instrument will contribute to its reliability and accuracy. The present study has shown that the new version of Preffi does produce reliable assessments for those project aspects covered by criteria that show enough variance. The decision study shows that drawing sufficiently accurate conclusions requires 4 to 12 assessors at the level of individual criteria, and least 5 at the level of clusters of criteria. At the level

of the project as a whole, sufficiently accurate assessment on the basis of all criteria can be done with only 2 assessors. These are encouraging findings for the development of a measurement instrument.

Nevertheless, we still need to try and reduce the number of assessors required for a reliable assessment, as the number of available assessors will often be limited in practical situations.

There are various ways to ensure more reliable assessment. One of the main sources of problems that contributed to the relatively low reliability found in the present study were the five-page project descriptions to which we asked our respondents to apply Preffi. This became clear from the numerous comments by correspondents stating that the descriptions offered too little information to adequately answer the questions asked in the operationalisations. In other cases, assessors had to conclude from a single sentence in the project description that a particular criterion had been met. It may be assumed that the reliability of the assessments will be greatly improved if the assessment is based not only on a project description or project plan, but also on a discussion with the project manager to obtain additional information. This is particularly true for aspects relating to the project management, support and commitment, and the question whether the right people have been approached or involved, which are difficult to assess purely on the basis of project descriptions. This idea of including discussions with colleagues, project managers or directors in the assessment procedure was frequently brought up in the interviews we had with our respondents. The Preffi development team has indeed found that when different assessors have assigned different scores to a project, discussions usually quickly result in consensus about the definitive score for a particular criterion.

Heuvelmans and Sanders suggested some other measures that can be taken to improve agreement between assessors (Heuvelmans and Sanders, 1993). One option that can be expected to lead to considerable improvement is to train users in applying the instrument. Another important factor is that the assessors should be made aware that their work will be checked by others, as they will then deviate less from the assessment framework and the instructions. In addition, Heuvelmans and Sanders emphasise the importance of unambiguous and manageable instructions for the assessors. Their proposals have already been incorporated in the recent, more extensive version of the User Manual referred to above (Molleman et al., 2003). Further studies of the reliability of the definitive version of Preffi 2.0 will have to show whether these adjustments have indeed improved the instrument's reliability.

The present study has yielded a number of recommendations for follow-up research. One aspect that needs to be ensured in future studies is that there is enough difference in quality between the projects to be assessed, so as to allow the G coefficients to be determined for individual criteria.

5.4.3 Expectations for the near future

Some of the assessors who contributed to the present study wondered whether the norms used in the present version of Preffi are not too strict, although the findings did not indicate that they were too strict (only that Preffi is not always easy to use). The decision to use strict norms had in fact been taken after extensive debates in the scientific advisory board and among the project team.

It is our expectation that the Preffi 2.0 criteria will gradually come to be used more and more, and will be referred to in project descriptions over the coming years. We think this is a realistic assumption, as the present study showed that the four projects assessed had high scores for the main criteria from Preffi 1.0, whereas scores for the criteria newly introduced in Preffi 2.0 were as yet lower. This implies that Preffi 2.0 also has an educational function.

Preffi 2.0 will only come to be extensively used if an active implementation programme is started, with the aim of consolidating Preffi's position as the standard in the field. It is the task of professional organisations, the inspectorate, umbrella organisations, support services, educational institutions and particularly executives to ensure that practitioners working in their organisations use Preffi as the standard quality assurance instrument.

Chapter 6 Project quality rating by experts and practitioners: experience with Preffi as a quality assessment instrument

6.1 Introduction

The Preffi is a Dutch quality assessment instrument for health promotion interventions, based on research findings about programme and project aspects affecting effectiveness and quality. Effectiveness is one of the important aspects of quality; other aspects usually distinguished include ethics and client satisfaction. Preffi was developed in the mid-1990s, after elaborate consultations with researchers and practitioners. In the early 2000s, it was thoroughly revised into a new version, called the Health Promotion Effect Management Instrument, the Preffi 2.0. The revision process was once again undertaken in close collaboration with researchers and practitioners, and was partly based on research into the instrument's usefulness and reliability. Both of these development processes have been described in detail elsewhere (Molleman et al., 1995; Molleman, 1999; Molleman et al., 2003; Peters et al., 2003). Preffi is the most elaborate quality assessment instrument in this domain and has received considerable attention both in the Netherlands and internationally (Keijsers and Saan, 1998; Molleman et al., 2004; Van den Broucke et al., 2004).

The instrument is mainly intended to allow practitioners – whether or not together with colleagues – to assess their own projects in terms of whether conditions for effectiveness are met and to suggest improvements. Although Preffi is mainly intended to help practitioners improve their own projects, we felt it was important to examine the reliability of using the instrument to assess their own as well as other people's projects.

The assessment aspect is more fully specified in Preffi 2.0 than in the 1.0 version, which did not clearly indicate when a project satisfied a particular quality criterion (Molleman and Hosman, 2003). To this end, elaborate operationalisations and norms were developed for the 2.0 version, which were expected to

provide users with clear instructions and thus make their assessments more reliable. A preliminary tests of the feasibility of a draft 2.0 version showed that users did indeed appreciate this specificity (Meurs, 2002; Molleman et al., 2004). The test also provided some indications of the instrument's reliability, as well as suggestions for improving the descriptions in Preffi, and allowed us to design a follow-up study into the reliability by providing indications for aspects like the number of assessors needed and the need to include a variety of projects, write a clear manual and train Preffi users. The present paper reports on the design and results of this follow-up study, concentrating on the reliability of assessing *other people's* health promotion (HP) projects with or without the use of Preffi, and on comparisons between the assessments made by practitioners using Preffi and those made by experts assessing the same projects.

The present study tried to answer a number of research questions. (1) Do quality assessments of HP projects by different experts show sufficient agreement to be used as a criterion to evaluate Preffi-based assessments of the same projects by practitioners? We hypothesised that experts, because of their experience as professional assessors on assessment committees, would share the same quality standards and hence show greater agreement in their assessment of projects (i.e., be more reliable) than practitioners. (2) Do the experts' greater experience and clearer standards lead to stricter, and therefore more negative judgements than those of the practitioners? (3) The Preffi criteria were deliberately made rather strict, with the aim of stimulating practitioners to adopt a critical attitude and to identify aspects of the projects that could be improved. This raised the question whether Preffi-based assessments by practitioners would be stricter than their intuitive assessments of the same projects. (4) Does the explicit definition of the Preffi norms make the practitioners' Preffi-based assessments more reliable than their intuitive assessments of the same projects? (5) How reliable are the Preffi-based assessments and how many assessors are required for a sufficiently reliable and accurate assessment?

Preffi 2.0 consists of 39 quality criteria (see for criteria table 4), subdivided into 8 clusters (problem analysis, determinants of behaviour and environment, target group, objectives, intervention development, implementation, evaluation and contextual conditions and feasibility). The names of the first 7 clusters show that Preffi 2.0 is based on planning models (Green and Kreuter, 1999; Bartholomew et al., 2001), whereas the final cluster emphasises the importance of project management. Each of the 39 criteria has been operationalised as one or more yes/no questions, which together yield a rating of 'weak', 'moderate' or 'strong' for that particular criterion. In addition, 12 of the criteria allow users to select the option 'not assessable'. The ratings for the various criteria in a clus-

ter are combined into an overall mark for the cluster, and finally into an overall mark between 1 and 10 for the project as a whole. (See also www.preffi.nl)

6.2 Methods

The present study examined the possible differences in quality assessments of 20 projects in three different conditions: (1) experts giving an intuitive judgement based on their expertise and experience, without any further guidelines; (2) practitioners giving an intuitive judgement and (3) the same practitioners assessing the same projects using Preffi 2.0. Before their Preffi-based assessment, the practitioners attended a one-day training session on using Preffi 2.0. After the results of the various assessments had been collected and analysed, they were discussed in a consensus meeting with the experts, the practitioners and the Preffi project team.

6.2.1 Study population

We decided to use three experts and three practitioners to assess the projects, as the preliminary test of the draft version of Preffi 2.0 had shown that this number would allow the reliability to be determined with sufficient accuracy.

The three experts we selected had for many years been professionally involved in the allocation of grants for research, development or implementation projects for health promotion. Their respective backgrounds were in research, policymaking and health promotion practice. The experts were familiar with Preffi 1.0 and knew that a new version was being developed; one of them was also familiar with the content of the new version.

Of the three practitioners, one was employed by a municipal health service, one by a mental health and addiction care institute and one by a national health education agency; they had had 6, 4 and 1 years of work experience, respectively. They were all university educated, with degrees in, respectively, medical biology, educational theory and health promotion. The practitioners had some knowledge of Preffi 1.0, and knew that Preffi 2.0 was being developed, but were unaware of its content until the training session.

6.2.2 Project selection

Various sources were used to collect 32 project descriptions: entries for the Preffi award, a national project database, and direct contacts with a number of

HP or prevention practitioners. The projects related to the various domains of HP practice in the Netherlands (general health, mental health, addiction, home care). All project descriptions had the same structure: background; choice of objectives, target groups and interventions; implementation; organisation; evaluation. Two members of the Preffi project team used Preffi 2.0 to assess all projects as good, moderate or poor. From these 32 projects, we ultimately selected 20, based on three criteria: (a) the availability of a sufficiently informative project description, allowing a large number of Preffi criteria to be scored; (b) sufficient coverage of the various domains and themes; and (c) sufficient numbers of good, moderate and poor projects, so that the objects to be assessed were representative of the total 'population' of projects and would yield enough variance in the scores between the projects to allow a realistic reliability assessment.

6.2.3 Measurement instruments

The practitioners and experts were invited to assess the projects by reading the 20 project descriptions, which were offered in a different order to each of the three assessors, and to fill in a scoring form for each project. This scoring form asked for intuitive marks between 1 and 10 to be given for eight general aspects of each project (problem analysis, determinants, target group, objective, intervention development, implementation, evaluation and contextual conditions) as well as an overall mark for the project as a whole. Since these aspects correspond to the clusters used in Preffi, they will be referred to below as clusters. In addition, the assessors were asked to indicate whether and to what extent they had already been familiar with each project and whether this had influenced their assessment in a negative or positive sense.

After they had attended a training session on the use of Preffi 2.0, the practitioners were invited to assess the 20 projects once more, this time using Preffi 2.0. This involved using a Preffi 2.0 scoring form for each project and applying the Preffi 2.0 operationalisations and norms. Once again, the projects were offered in different order to the three participants.

6.2.4 Data analysis

The reliability of the intuitive judgements by the experts and practitioners and the Preffi-based assessments by the practitioners was analysed using the generalisability theory (Shavelson and Webb, 1991). The generalisability theory tries to assess the influence of various sources of error in a measurement. In an ideal

situation, the differences in scores can be attributed to differences between the objects to be assessed, in this case the projects, but they may also be caused by various error sources, like different views of the assessors.

The varcomp procedure in SPSS was used to determine the magnitude of the various sources of variance. This was then used to calculate the generalisability coefficient (G) and the Standard Error of Measurement (SEM). G is a measure of reliability (minimum acceptable value .70), while SEM is a measure of accuracy (maximum acceptable value 0.26). Subsequently, we estimated the values for G and SEM for a single assessor and calculated the minimum number of assessors needed to achieve these minimum and maximum values.

A multi/group regression analysis in Lisrel 8.20 was used to test whether the assessors used the same weights in assessing various aspects of the projects.

6.3 Results

6.3.1 Reliability and accuracy of the intuitive assessments

Combining all intuitive cluster assessments by the three experts yielded acceptable values for the generalizability coefficient and the SEM (see table 1). The minimum value for G (.70) and the maximum value for SEM score (.26) were not achieved using one assessor (G=.65; SEM=.45).

At the level of individual clusters, the experts didn't obtain any G values above the minimum (.70) or SEM values below .26 for any of the aspects.

The agreement between the experts' assessments was smaller than that between the intuitive assessments by the practitioners and that between the practitioners' Preffi-based scores at criterion level. The experts also had higher SEM scores (i.e., were less accurate) than the practitioners for nearly all clusters. This was not in line with our expectations.

The lack of agreement between the experts was also evident from the Lisrel regression analysis. The regression weights of the individual assessment aspects differed significantly between the experts ($\chi^2 = 55.92$, $df = 16$, $p = .000$). Table 2 shows the individual regression weights for the intuitive scores of the different experts for each of the eight clusters, relative to the overall scores for a project. For all experts, the assessment of the intervention is a major predictor of the overall score, whereas the degree to which aspects such as determinants, objec-

	experts, intuitive		practitioners, intuitive		practitioners' scores, using Preffi cluster scores		criteria scores**	
	G	SEM	G	SEM	G	SEM	G	SEM*
<i>values for all clusters together **</i>	0.83	0.26	0.98	0.18	0.85	0.49	0.96	0.17
<i>individual clusters</i>								
problem analysis	0.51	0.84	0.84	0.68	0.76	0.86	0.82	0.6
determinants	0.66	0.71	0.89	0.69	0.79	0.87	0.82	0.57
target group	0.43	0.74	0.83	0.71	0.65	0.81	0.65	0.45
targets	0.16	0.96	0.82	0.65	0.67	1.01	0.79	0.6
intervention development	0.08	0.98	0.67	0.69	0.65	0.76	0.94	0.18
implementation	0.52	0.76	0.7	0.79	0.55	0.94	0.89	0.36
evaluation	0.35	1.1	0.85	0.83	0.79	0.88	0.95	0.33
contextual conditions	0.38	0.7	0.61	1.03	0.69	0.94	0.84	0.45
overall project score	0.36	0.81	0.64	0.57	0.67	0.86		

* corrected for 10-point scale

** clusters and criteria scores are considered fixed facets.

minimum value for G: .70; maximum value for SEM: .26; values within these limits are shown against a shaded background

Table 1: G coefficient and SEM values using three assessors

tive, target group, implementation and contextual conditions played a role differed for the three experts.

The practitioners also differed in term of regression weights ($\chi^2=44.12$, $df=16$, $p=.000$). An interesting aspect is that, in addition to the intervention, they appear to have attached greater weight to determinants, evaluation and contextual conditions in their overall assessment.

6.3.2 Strictness of assessments: level of scores

Our expectations about the level of the scores proved partly correct. Across all projects, the experts gave an average score (on a scale of 1 to 10) of 5.6, against a score of 6.2 by the practitioners, a difference which was significant ($p=.004$) (See table 3).

assessor	problem analysis	deter- minants	target group	objec- tives	interven- tion	implemen- tation	evalua- tion	contextual conditions
expert 1	0.11	-0.09	0.27	0.21	0.29	0.18	0.07	0.07
expert 2	0.12	0.43*	0.16	-0.09	0.22*	0.06	0.1	0.51*
expert 3	0.11	0.17	0.13	0.24*	0.25*	0.16*	0.07	-0.05
practitioner 1	-0.07	0.66*	0.14	-0.34*	0.16	0.11	0.40*	0.2
practitioner 2	0.15	0.05	0.03	0.35*	0.33*	0.22*	0.11	0.21*
practitioner 3	0.14	0.29	0.28*	-0.13	0.33*	-0.09	0.29*	0.21*
sum for all assessors	0.09	0.25	0.17	0.04	0.26	0.11	0.17	0.19
sum for experts	0.11	0.17	0.19	0.12	0.25	0.13	0.08	0.18
sum for practitioners	0.07	0.33	0.15	-0.04	0.27	0.08	0.27	0.21

* significant

Table 2: regression weights per assessor for the aspects assessed, relative to the overall score of a project

	experts, intuitive		practitioners, intuitive		practitioners, using Preffi	
	X	SD	X	SD	X	SD
overall project score	5.6*	1.5	6.2	1.6	6.2	1.8
problem analysis	6.2	1.7	6.5	1.9	5.7	2.1
determinants	5.6	1.6	5.8	2.2	5.4	2.2
target group	5.5	1.4	6.0	1.9	5.9	1.7
targets	5.7	1.7	5.8	1.8	5.9	2.1
intervention development	5.7 *	1.7	6.6	1.5	7.0	1.6
implementation	5.9	1.5	6.3	1.8	6.4	1.9
evaluation	5.0	2.0	5.5	2.4	5.3	2.2
contextual conditions	6.0	1.3	5.4	2.2	5.8	2.0

* p < .01

Table 3: average overall and cluster scores (range 1-10) for each of the test conditions

Experts assigned lower scores for all of the individual aspects, except contextual conditions. The differences were, however, only significant for the aspect of intervention ($p=.001$).

The average score for the overall judgement of a project by the practitioners was just as high when they were asked to assess the project using Preffi as when they were asked to give an intuitive assessment. Some of the individual cluster scores were higher using Preffi, and some were lower, but the average score was the same with or without Preffi. Whereas two of the practitioners gave a less favourable average assessment with Preffi, the third gave a more favourable assessment, even though she thought she had been stricter in her judgement.

6.3.3 Reliability of Preffi-based assessments compared to intuitive assessments

We present two measures of Preffi scores: scores on clusters and the 'Preffi criteria scores'. For the latter values for G and SEM were calculated per cluster on the basis of the scores for the various criteria within that cluster.

The highest reliability (G) and accuracy (SEM) values for the overall cluster scores were those for the intuitive assessments by the practitioners and the Preffi-based assessments (table 1). These two types of assessment even produced acceptable G and SEM values with only one assessor. The practitioners also produced acceptable G values for most of the individual clusters, whether they were using intuitive assessment or the Preffi criteria scores. The greatest accuracy of scores was achieved by using the Preffi criteria scores. This is not surprising, since the SEM is determined by multiple assessments, which means that test lengthening plays a role.

We were surprised by the low reliability (G values) and low accuracy (high SEM values) we found for most of the cluster scores using Preffi.

6.3.4 Reliability and accuracy of Preffi-based assessments

Table 4 shows the reliability and accuracy of the assessments by the three practitioners across the 20 projects. It shows the relative size of the variance estimates for each source of variance (project, assessor and interaction effect between project and assessor) for each criterion and cluster and for Preffi as a whole. 'Project' as a source of variance is an estimate of the true score variance,

	number of assessors	G abs SEM		% of total variance			SEM < .26 for n=	then G abs:
		3	3	project	assessor	pro*ass		
2 Problem analysis	82	20		43.3%	15.8%	40.9%		
2.1 Nature, severity, scale of problem	58	35		31.2%	9.4%	59.4%	6	.73
2.2 Distribution of problem	63	39		36.5%	14.2%	49.3%	7	.80
2.3 Problem perception by stakeholders	62	42		35.0%	10.6%	54.4%	8	.81
3 Determinants	82	19		44.7%	21.1%	34.2%		
3.1 Theoretical model	52	36		26.3%	22.6%	51.1%	6	.68
3.2 Contributions of determinants	43	33		19.8%	19.2%	61.0%	5	.55
3.3 Amenability of determinants to change	58	35		31.4%	2.5%	66.2%	6	.73
3.4 Priorities and selection	61	45		34.0%	7.3%	58.7%	9	.82
4 Target group	65	15		20.1%	7.8%	72.1%		
4.1 General and demographic characteristics	43	40		20.3%	24.2%	55.6%	8	.67
4.2 Motivation and possibilities of target group	51	35		25.7%	0.0%	74.3%	6	.67
4.3 Accessibility of target group	00	28		0.0%	0.0%	100.0%	4	.00
5 Objectives	79	20		36.7%	17.4%	46.3%		
5.1 Objectives fit in with analysis	36	41		15.6%	28.0%	56.4%	8	.60
5.2 Objectives specific, specified in time, measurable	81	32		58.8%	8.8%	32.4%	5	.88
5.3 Objectives are acceptable	20	38		7.8%	0.0%	92.2%	7	.37
5.4 Objectives are feasible	50	48		24.9%	13.7%	61.4%	11	.79
6 Intervention development	94	06		40.1%	2.1%	57.8%		
6.1a Fit strategies/methods to target, target groups	00	33		0.0%	3.1%	96.9%	5	.00
6.1b Previous experiences with intervention(s)	75	33		50.1%	0.0%	49.9%	5	.83
6.2a Duration and intensity of intervention	70	38		44.1%	9.8%	46.2%	7	.85
6.2b Timing of intervention	36	38		15.6%	33.2%	51.1%	7	.56
6.3a Participation of target group	75	32		50.3%	0.0%	49.7%	5	.84
6.3b Fitting to 'culture'	53	30		27.4%	7.8%	64.8%	4	.60
6.4 Effective techniques (recommended)	34	27		14.8%	33.1%	52.1%	4	.41
6.5a Fitting to intermediary target groups	51	33		25.5%	0.0%	74.5%	5	.63
6.5b Characteristics of implementability intervention(s)	15	38		5.5%	1.0%	93.5%	7	.29
6.6 Coherence of interventions/activities	34	51		14.8%	0.6%	84.6%	12	.68
6.7 Pretest	41	38		18.9%	5.7%	75.4%	7	.62
7 Implementation	89	12		41.2%	4.7%	54.1%		
7.1a Mode of implementation: top-down bottom-up	42	39		19.5%	10.4%	70.2%	7	.63
7.1b Fitting the implementation interventions to intermediaries	64	28		37.4%	9.4%	53.2%	4	.71
7.1c Appropriateness of supplier for intermediaries	62	28		35.4%	3.2%	61.4%	4	.69
7.2 Monitoring and generating feedback	39	49		17.8%	11.7%	70.5%	11	.70
7.3 Incorporation in existing structure	42	42		19.1%	11.3%	69.5%	8	.66
8 Evaluation	95	11		55.5%	0.9%	43.6%		
8.1 Clarity and agreement on principles of evaluation	76	34		51.7%	0.6%	47.7%	6	.87
8.2 Process evaluation	91	22		76.6%	4.2%	19.2%	3	.91
8.3a Has a change been measured?	41	41		18.8%	0.0%	81.2%	8	.65
8.3b Was the change caused by the intervention?	62	32		35.0%	5.7%	59.3%	5	.73
8.4 Feedback to stakeholders	29	36		12.2%	0.0%	87.8%	6	.45
1 Contextual conditions and feasibility	84	15		45.7%	15.6%	38.7%		
1.1 Support/commitment	75	31		50.5%	7.1%	42.5%	5	.84
1.2 Capacity	74	37		48.6%	20.1%	31.3%	7	.87
1.3a Expertise and characteristics of project manager	00	27		0.0%	0.0%	100.0%	4	.00
1.3b Focal points for leadership	00	37		0.0%	26.3%	73.7%	6	.00
Total project (values for all criteria together)	.96	.06		51.1%	5.9%	43.0%		
average number of assessors required:							6,36	

Table 4: G and SEM values for each individual item, per cluster, and for the assessments as a whole, plus the numbers of assessors needed to achieve an accurate assessment, per item, with the corresponding G values

which ideally should be as high as possible. The projects showed large enough differences for most criteria. Four of the five criteria which produced little or no variance between the projects were frequently given a 'not assessable' score. The variance attributable to differences in views between the assessors should be as low as possible. A number of factors yielded low values in this respect, such as the criteria of 'theoretical model', 'objectives fitting in with the analysis', 'timing of the intervention' and 'effective techniques', which may indicate that differences in the amount of experience might play a role in assessing these aspects.

The G and SEM values in this table for Preffi 2.0 as a whole and at criterion level have also been included in the column headed Preffi criteria scores in table 1. A combined analysis for the criteria per cluster yielded acceptable reliability (G) and accuracy (SEM) values for all clusters except that for target group, indicating that, at this level, Preffi is sufficiently reliable and accurate using three assessors.

Table 4 also shows for each criterion the number of assessors needed to obtain an acceptable SEM, that is, a sufficiently accurate assessment. This number ranges from 4 to 12 for the various criteria, with an average of 6.36 assessors per criterion.

6.3.5 Notable aspects of project scoring

he time required to assess a project using Preffi was initially about twice as long as for an intuitive assessment of the various aspects of a project. After the assessors has assessed about 10 projects with Preffi, this difference in time required between intuitive and Preffi-based assessment had disappeared, with both types of assessment requiring an average of 50 minutes per project.

Prior knowledge about a project could have a clear impact on the assessment scoring. The respondents indicated that prior knowledge played a major role in the project score in 10% of all projects. In half of these cases, prior knowledge led to a higher score, and in the other half to a lower score. The consensus meeting confirmed that some of the striking deviations in scores were based on this phenomenon.

The practitioners differed in the extent to which they used the 'not assessable' option in assessing projects with Preffi. This option can be used for instance on items relating to aspects of project management and implementation and evaluation. These aspects are often difficult to assess on the basis of project descriptions. Although the training sessions had paid considerable attention to scoring these

aspects, the three assessors turned out to have different views of their task in this respect when actually assessing the projects. One of the factors affecting these views appears to be the amount of experience, as assessors with greater experience used the ‘not assessable’ option less frequently than those with less experience.

Preffi suggests a scoring method to calculate cluster scores from criteria scores. However, the practitioners were allowed to deviate from this algorithm by attaching greater weights to certain criteria than to others. Figure 1 shows that the practitioners did indeed make use of this opportunity: P1's Preffi-based assessment shows lower scores for most clusters than would have been justified by the criteria scores, while P3's scores were higher than justified by the criteria scores.

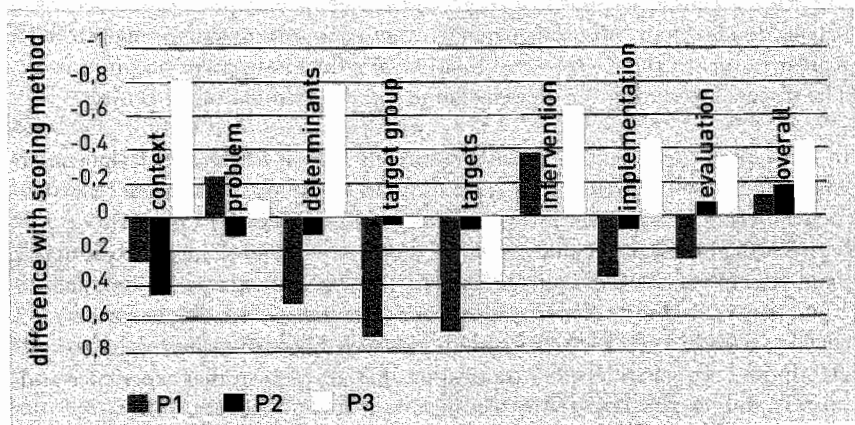


Figure 1: differences between average calculated sum scores according to scoring method and actual sum scores for three practitioners for each of the clusters and for the project as a whole

6.3.6 Consensus meeting

The consensus meeting was used to discuss the assessments and the participants' experiences. They exchanged arguments for the various scores, with the intention of achieving consensus. The following experiences and insights emerged. The participants had often found it hard to assess a project on the basis of a project description alone. They would have preferred to talk with the project leader to get more information on aspects not included in the project description. They also reported that having prior knowledge about a project made it easier to assess it.

The participants were aware that people have their own assessment strategies and may attach greater importance to certain aspects than others, often based on their own profession, background and experience. One practitioner may feel more strictly bound to use the Preffi operationalisations than another, who may prefer to include the project's intentions in his or her assessment.

One of the experts emphasised that assessment processes in groups tend to seek consensus on the better projects (Vries, 1993). Comments on shortcomings in a project quickly result in the project being dropped. Consensus on poor projects was indeed rapidly achieved in the discussions.

The interpersonal differences we found were deemed to be unavoidable and were even regarded as a positive aspect, since they may help the group achieve a balanced judgement. All participants felt that consensus meetings ought to be a regular feature of the process used to arrive at a final judgement on projects. The discussions during the meeting led to rapid agreement on all projects.

In most cases, the other participants eventually tended to agree with the expert whose assessments differed most from those of the group. This expert's arguments focused on applying the right theory and/or achieving a good fit between the intervention and its intended objective. Even though the participants felt that it was important to use various perspectives, content-oriented arguments proved most decisive at this meeting. It was clear that less experienced practitioners had in some cases been swayed by well-written project descriptions. Research in other occupations has also shown that experience and expert opinions play important roles in assessments (Wagenaar, 1987).

The meeting clearly showed that assessing projects is not a simple matter, and one involving many aspects. Nevertheless, there was a clear consensus that Preffi helps to identify weak aspects of a project and indicates points that could be improved.

6.4 Discussion and conclusions

This study assumed that expert assessments could be used as an external criterion with which the reliability of Preffi-based assessments could be compared. The findings show that, as far as experts are concerned, at least three assessors are needed to achieve reliable and accurate assessments. Agreement between the experts was weaker than between the practitioners. The intuitive assessments by practitioners and those based on Preffi criteria produced considerably more reliable and accurate project assessments. These three assessors together produced

virtually the same high G and low SEM values, which more than met the minimum values defined in advance. The practitioners preferred the Preffi-based assessment, as it produced a more balanced assessment and provided them with a clear instrument to assess and improve projects.

The study showed that project assessment involves a multitude of aspects, some of which are revealed by the analyses of the scores. Our regression analyses revealed differences among both experts and practitioners in the weights they attach to various aspects in their overall judgement. In addition, they showed that the practitioners attached different weights to the various Preffi criteria that have to be combined in a cluster score. Aspects like prior knowledge about a project also play a major part in project assessments, and could have a positive or a negative influence.

The respondents found it hard to assess a project purely on the basis of a project description, as descriptions often lack certain information. Preffi criteria requiring an assessment of project management aspects, such as the appropriateness of the supplier (7.1.c), adequate support (1.1), capacity (1.2), acceptable and realistic objectives (5.3, 5.4), tailoring to the 'culture' of the target group (6.3b) and practical feasibility (6.5) are particularly hard to assess. This finding fits in with those of an earlier study of the draft version of Preffi 2.0. Hence, we would recommend including a discussion with the project leader in the assessment procedure.

The consensus meeting revealed the importance of a discussion between assessors to achieve consensus and a final assessment. Consensus was achieved without great problems, and the participants recognised that a project could be looked at from different perspectives. This should not be regarded as a 'source of error', as the generalisability theory would have it: the participants regarded different perspectives among assessors not only as useful, but even as essential. The meeting showed how complex an adequate assessment can be. The expert whose opinions deviated most from those of the others actually managed to sway most of the others. This also indicated that Preffi includes a number of items (e.g. theory (3.1), determinants (3.2), the fit between analysis, objective and intervention (5.1 and 6.1a) and coherence (6.6)) that require more than an assessment by practitioners. They also require an expert opinion to assess whether the justifications given are correct; the least experienced practitioner found it hard to see through the favourable projects descriptions.

As expected, the experts gave stricter assessments of projects than practitioners. Our hypothesis that the use of the strictly worded Preffi criteria would result in

practitioners also giving more negative judgements than in their intuitive assessments was not confirmed. Although the practitioners themselves thought they had been stricter when using Preffi, the empirical data did not bear this out.

Our findings show that the practitioners' cluster scores based on the Preffi criteria resulted in a low G value and a high SEM value, which were much poorer than their intuitive assessments of the same aspects. A possible explanation is that Preffi induces people to examine certain aspects more specifically, leading to more differentiated assessments with greater variation. This effect is further increased by the different weights attached to the criteria scores used to calculate cluster scores, as figure 1 shows. This means that the calculation of cluster scores from criteria scores does not contribute to the instrument's reliability.

An important underlying question of the present study is that of the reliability of assessments based on Preffi 2.0. As regards the overall assessment of a project, two assessors should be enough to produce a sufficiently reliable and accurate judgement, while three assessors would be required at cluster level, and between four and twelve assessors (average 6.36) would be needed at criterion level.

A sufficiently accurate assessment of a criterion will lead to an acceptable G coefficient ($> .70$) for 17 of the 39 criteria. Although this is not yet enough, the findings of the present study are better than those for the draft version. The present version includes 29 criteria that require fewer assessors to achieve an SEM $< .26$ than the draft version. The average number of assessors needed per criterion has dropped from 8.13 for the draft version to 6.36 for the present version. This may have been caused by improvements made to the draft version, (since all revised criteria produced better scores) or by the training provided to the assessors, or by the fact that the present study used a larger number of projects, allowing the respondents more practice with the instrument.

Our overall conclusion is that the reliability and accuracy of Preffi 2.0 are not yet sufficient to allow a project to be assessed by one assessor. In fact, the participants of the consensus meeting questioned whether this would ultimately be desirable.

Certain improvements can be introduced to increase Preffi's reliability. Firstly, assessors should discuss their assessments in a consensus meeting. Secondly, a consultation with the project leader should be included in the assessment procedure in order to supply information that was lacking in the project description, especially on project management aspects. Finally, adequate scoring of some Preffi criteria requires an expert's opinion. This aspect cannot be ade-

quately addressed in a list of general guidelines like Preffi; it means that the general insights produced by Preffi will have to be specified for certain themes. Such studies can then indicate the relevant theories for the specific domain, the suitable interventions and the specific success and failure factors. In the future, expert systems on the Internet, linked to a digital version of Preffi, could provide this.

Further research will have to show whether this will actually improve the reliability of the Preffi instrument.

1. The first step is to identify the problem or question that needs to be answered. This involves understanding the context and the specific requirements of the task.

2. The second step is to gather relevant information and data. This can be done through research, consultation with experts, or by analyzing existing data sets.

3. The third step is to develop a plan or strategy to address the problem. This involves breaking down the problem into smaller, manageable parts and determining the best approach to solve each part.

4. The fourth step is to implement the plan and monitor progress. This involves carrying out the tasks outlined in the plan and keeping track of the results to ensure that the goal is being met.

5. The fifth step is to evaluate the results and make adjustments as needed. This involves comparing the actual results with the expected outcomes and identifying any areas where the plan needs to be revised.

Chapter 7 Conclusions and future prospects

7.1 Introduction

This thesis has discussed the development of Preffi 2.0 and studies into its validity, usefulness and reliability. Preffi 2.0 is an effect management instrument that intends to help health promotion specialists to increase the effectiveness of their interventions by applying the Preffi criteria and guidelines to these interventions. These criteria reflect both research findings on effect predictors and practitioners' experience with effectiveness, for instance relating to context and project management. The instrument can be applied to individual health promotion interventions or to projects involving multiple interventions or programmes including multiple projects.

Preffi is part of a wider national strategy to increase the quality and effectiveness of health promotion and disease prevention in the Netherlands. In the context of this strategy, various organisations, such as the Netherlands Institute for Health Promotion and Disease Prevention (NIGZ), the Trimbos Institute, the Netherlands Organisation for Health Research and Development (ZonMw), the Netherlands School of Public and Occupational Health (NSPOH), are working on a range of activities, like setting up databases with existing and model projects, analysing existing projects, training and assisting practitioners to help them improve their projects or programmes, and promoting the use of quality assurance policies in various organisations.

Besides Preffi, health promotion specialists also use other quality assurance instruments, especially Intervention Mapping (Bartholomew et al., 2001). Both instruments use the same principle of a systematic approach, and many of the Preffi criteria are also included in Intervention Mapping.

Intervention Mapping emphasises health education based on theory and research, and the adequate use of theory to develop the right interventions. This aspect is more fully and thoroughly worked out in Intervention Mapping than in Preffi, which may relate to the fact that Intervention Mapping was developed by researchers. By contrast, Preffi pays more attention to aspects of

practice and the usefulness of the instrument for practitioners. The differences between the two instruments are thus a matter of emphasis rather than of fundamental issues.

It is our intention to update Preffi regularly, thus creating a 'learning system' for the dynamic relation between research and practice.

Various strategies can be used to increase the effectiveness of programmes or interventions. One strategy that is commonly used – particularly at US universities – is to develop and test model programmes. Programmes that are found to be effective are then made available to local health promotion specialists, who are expected to apply them as accurately as possible. However, the process of systematically developing and evaluating such a programme may well take 10 to 15 years. A further problem is that large-scale implementation of such model programmes usually means that they have to be adapted to local situations and contexts. This strategy has been adopted by many other countries. In the Netherlands, it actually formed the basis of the 1997 – 2002 ZonMw prevention programme.

Another strategy is to stimulate local health promotion specialists to apply general principles and guidelines on effectiveness in their day-to-day routine. This strategy is particularly suitable for the Dutch situation, in which about 1200 local health promotion specialists are responsible for the development, application, implementation and improvement of health promotion programmes (Poel et al., 2005). These include not only model programmes but also programmes developed by the practitioners themselves for their particular situation. It is within this context that Preffi is a useful effect management instrument.

Our view is that a combination of the above strategies is required to improve effectiveness in practice.

This thesis describes the development of an effect management instrument intended to support the second of the above strategies, concentrating on the development of Preffi 2.0, and studies undertaken in the context of this development. This programme was started in 2000 and concluded with the launching of Preffi 2.0 in 2003.

This new version builds on the experience gained with the first version, which was developed in 1994 – 1995 in close consultation with Dutch health promotion specialists, and on its systematic implementation in the 1997 – 1999 period.

The questions this thesis has tried to answer are:

1. What lessons does the Preffi 1.0 programme teach us about the content, format and use of the instrument?
2. How was Preffi 2.0 developed and how were validity requirements met?
3. What can be said about the usefulness and reliability of Preffi 2.0?

This final chapter briefly reviews these questions and provides some suggestions for the future.

7.2. Lessons learned from the Preffi 1.0 programme in terms of the instrument's content, format and use

Chapters 2 and 3 of the present thesis discussed the lessons we learned from the Preffi 1.0 programme. Chapter 2 presented Preffi 1.0 and explained its format and content. The items, or effect predictors, in that version basically formed a checklist of items judged to be relevant to improving effectiveness and quality, and had not been operationalised. In a study of the systematic implementation programme for Preffi 1.0, we asked a representative sample of Dutch health promotion specialists to complete a questionnaire in 1997 (n=120) and 1999 (n=316). The study yielded information on the way health promotion specialists actually addressed the various effect predictors included in Preffi 1.0. It also assessed how many health promotion specialists were aware of the existence of Preffi, whether and how they were using it and what they thought about it in general.

In addition to the general national implementation activities, the implementation programme included a more intensified 5-day training programme, offered to a particularly motivated group of health promotion specialists (n=64). The study, which used a pre- and post-test design, showed that health promotion specialists are very much action-oriented and that their choice of interventions for their projects is made rapidly and is mainly based on pragmatic arguments. The health promotion specialists often find it difficult to motivate their choices on the basis of theoretical considerations. In addition they frequently have little idea of determinants and tend to formulate objectives in vague or inadequate terms. The same impression was gained from training sessions and the consultative discussions we had with other health promotion specialists.

Preffi 1.0 was perceived as useful and helpful and as meeting a need in terms of the above aspects. Preffi users reported that the instrument stimulated them to take a more systematic approach to their projects and helped them to address

determinants, theory and a careful definition of objectives and target groups, and hence to choose an intervention that suits the objectives. The respondents reported that they used Preffi in a flexible manner, in the sense of using only those parts of the instrument that are relevant to their own specific context. They also said that they did not always find it easy to use Preffi 1.0. The main problem was deciding when it could be said that a project satisfied a particular Preffi criterion. Since the effect predictors in Preffi 1.0 had not been operationalised, there was no standard against which they could assess their own work.

The various assessments in the implementation study yielded useful empirical data on the effect of the implementation programme. It turned out that the nationwide activities had contributed to greater awareness of Preffi's existence and to a favourable attitude towards the instrument. However, this had not led to a significant increase in its use. The latter was found to require a more intensive training programme. Another unexpected finding was that the health promotion specialists' superiors had hardly stimulated the use of Preffi as a quality assurance instrument. Apparently, managers of organisations assume that the professionals themselves will ensure that their work is of sufficient quality.

The Preffi 1.0 programme provided valuable input for the development of the new and improved version, Preffi 2.0. It showed that the instrument's *format* needed to reflect the iterative and cyclical nature of health promotion projects, and that each criterion had to be operationalised and provided with a norm. The instrument's *content* needed to incorporate not only various new research findings about the effect predictors included in Preffi 1.0, but also new aspects of implementation strategies, context variables, theoretical principles, the motivation of the target group, participation by the target group and the principles used in community projects.

As regards the instrument's *role and status*, there was a need to focus not only on the health promotion specialists themselves, but to achieve greater support from researchers and institutional management. The instrument's name needed to emphasise that it intended to evaluate conditions for effectiveness. To this end, Preffi's original Dutch name, which literally translates as Prevention Effectiveness Instrument was changed to Health Promotion Effect Management Instrument.

A number of critical comments can be made on the Preffi 1.0 programme. The studies were based on written questionnaires and information that was systematically gathered during the training programmes on the use of Preffi 1.0 or its

components. The data largely consisted of self-reports by users and non-users of Preffi. Their answers showed that health promotion specialists did indeed regard the Preffi criteria as relevant – a conclusion which was confirmed by our experiences during the training courses. Those who took part in the intensive training programme indicated that they were now able to apply the Preffi criteria more adequately, which in their opinion would lead to better projects. Whether this is indeed the case, and whether the quality of projects had indeed improved, was beyond the scope of our study. Additional research should show whether existing project plans will actually be adjusted, and whether health promotion specialists will actually change their professional attitude after they have been trained in the use of Preffi and have started using it.

The model of the implementation process that involved dividing it into a number of stages met our expectations. Our critical reflection on the implementation programme yielded two valuable new insights. The first was that the implementation plan, though carefully designed, had been largely aimed at the health promotion specialists themselves. The plan defined different objectives and corresponding interventions for health promotion specialists in the different stages within the implementation process. Although the implementation plan did recognise that institutional management, the professional association, researchers and trainers could provide important stimuli for health promotion specialists to start using an instrument like Preffi, it did not include any specific activities aimed at them. In the end, concrete implementation activities were only deployed towards the professional association. Ideally, the implementation plan should also have included activities aimed at researchers and institutional managers, which would have had to be tailored to the main determinants that influence whether such people will stimulate the use of Preffi. Such activities were never undertaken, however. This was partly because of a lack of capacity, but also because such activities would have required the recruitment of other people than the Preffi team members, as the latter were primarily focused on health promotion practitioners. Managers need to be addressed at their own level, using arguments tailored to them. This aspect of implementation, the strategy of using the right persons to address and convince the various stakeholders, is often neglected in health promotion.

The second insight gained from the implementation process was that implementing an instrument like Preffi requires long-term investments, which can only be achieved if there is an 'owner', that is, a party that feels responsible for the implementation, works on it consistently and is willing to invest if no external funding can be obtained for the project. This aspect should be more actively considered in the process of thinking about the implementation and

structural embedding of innovations; it should also have consequences for the resources and subsidies provided. It must be said that the idea of a committed 'owner' as an absolute precondition for a long-term implementation programme clashes with the current situation in the Netherlands, in which many organisations largely depend on short-term project funding for their implementation projects.

Preffi and the materials and activities to support its use have from the first been developed in close collaboration with the target population, the health promotion specialists. This has allowed us to clearly identify the consequences of proposals and assess what is possible or impossible to achieve in practice. This collaboration with the intended users has frequently yielded unexpected and useful solutions.

The downside of such a close collaboration with the health promotion specialists is that it becomes more difficult to keep the necessary distance to the target group. In the first few years of the Preffi programme, this was particularly evident in the tension between the two functions that Preffi was intended to fulfil, that of assessment and that of improvement, the focus being on the second aspect. Analyses of projects that took place, for instance, in the context of the intake workshops with the group of highly motivated health promotion specialists, often revealed that they failed to meet many of the requirements of good health promotion. This then creates the problem of whether knowledge of the contextual conditions in which the health promotion specialists have to work should determine one's tolerance of such shortcomings. How strict can one's judgement be without jeopardising one's close ties with the health promotion specialists? This is a delicate balancing act, which has often resulted in the decision to emphasise the potential for improvements and the provision of professional support, although the assessment aspect was not neglected (Molleman, 1999).

7.3 Preffi 2.0: development and validity

Chapter 4 described the systematic process of developing Preffi 2.0. This process consisted of seven steps which together determined the instrument's content or construct validity, in a careful and wide-ranging consensus process.

In the first two steps, we determined Preffi's position within the wider framework of quality management and decided to continue our systematic approach as the basic structuring principle for Preffi. This resulted in 5 dimensions

(analysis, development, implementation, evaluation and contextual conditions), for each of which we formulated a set of relevant effect predictors.

To this end, steps 3 and 4 involved a screening, by the Preffi project team and an advisory committee of researchers (WAR), of the effect predictors already used in Preffi 1.0, and suggested additional effect predictors, using five selection criteria: *relevance*, *evidence from research*, *generalisability*, *modifiability* and *measurability*. The screening process showed that the selection criteria of *generalisability*, *modifiability* and *measurability* were relevant and were met by all effect predictors used in Preffi 1.0.

The *relevance* of an effect predictor to effectiveness was deemed the most important selection criterion; it refers to the proven or assumed influence that a particular characteristic of a project has on the project's effectiveness. The scores for *evidence from research* are an indication of the 'evidence-based' nature of the effect predictors (or criteria) used in Preffi.

Steps 5 and 6 involved describing the importance of each effect predictor in Preffi in a document called Explanatory Guide. In addition, each effect predictor (criterion) was operationalised and provided with a norm. After several rounds of consultations with the WAR and a practitioners' advisory committee (PAR) we produced a draft version of Preffi 2.0. In step 7, this draft version was tested for usefulness by 35 experienced practitioners. Their comments were used to produce the definitive version of Preffi 2.0.

This careful systematic procedure ensured the instrument's content validity. The WAR and PAR committees played an important role in this. The WAR included representatives of all major approaches used in Dutch health promotion research, while the PAR was composed of a large sample of health promotion specialists from various sectors and with different operational styles.

A possible criticism could be that we used a rather broad definition of the concept of evidence in our selection criterion *evidence from research*. An important reason for this was that the predictive value of some effect predictors which are deemed important is hard to prove by controlled quantitative research (especially randomised controlled trials), which is usually seen as the highest form of scientific evidence (Cochrane Collaboration; CSAP (Center for Substance Abuse Prevention); NHS Centre for Reviews and Dissemination (CRD); Rada et al., 1999). For instance, the principles of a systematic approach are often based on rational argument, consensus and experience, as well as the results of repeated case studies. There is as yet no specific, internationally accepted hierar-

chy of types of evidence (McQueen, 2000; McQueen and Anderson, 2002) and there is increasing support for the view that other methods of proof besides randomised controlled trials are also legitimate (Tones, 1999; McQueen, 2000; Koelen et al., 2001; Saan and Haes, 2005).

The Explanatory Guide (provided on the CD-ROM) includes scores for each of the effect predictors, relating to the hierarchy of evidence we used, with an explanation of the scores. This shows that 20 of the effect predictors attained a score of 2, which means that evidence is available from practical experience and theoretical assumptions, but no empirical evidence. Empirical evidence from meta-analyses or systematic reviews was available for 16 of the effect predictors and for all 9 effective techniques (Preffi item 6.4), giving them a score of 5. Three of the effect predictors were given a score of 2 – 5, since there was wide consensus about the predictor from practical experience and theoretical assumptions, but the available empirical evidence from meta-analyses related only to part of the predictor. Our overview (page 4 of the Explanatory Guide) indicates that there may well be further improvements in the available evidence over the coming years; it reflects the current scientific status of the health promotion field. We have to remain cautious, since we based our decisions on a literature study which was unavoidably limited by time constraints. There may well have been studies which we have failed to find, and other areas of research may produce evidence for effect predictors that could also be applied to and be valid for our health promotion. Since we have worked in as transparent a way as possible, our overview could be a useful point of departure for further research that could gradually extend the evidence as new research data become available. This results in an interesting research agenda for the coming years. Since research into health promotion and disease prevention is rapidly developing, it is to be expected that a new, updated version of Preffi will be needed within a few years.

The content validity of Preffi 2.0 was recently confirmed by a study in the context of the European project called Getting Evidence into Practice, in which national health promotion agencies from the various EU countries try to develop a joint European Quality Assessment Tool. Preffi was found to compare very favourably with all other quality assessment tools in the world. Other tools included very few, if any, effect management criteria that are not included in Preffi, whereas many Preffi items not included in other instruments were deemed relevant by an international panel of experts in the Delphi rounds (Vermeulen et al., 2005).

7.4 Preffi 2.0: usefulness and reliability

Chapters 5 and 6 reported on two studies to test the usefulness and reliability of Preffi 2.0. In the first study, 35 practitioners used the draft version to each assess two projects (from a total of four). The respondents commented favourably on the usefulness of Preffi 2.0. The draft version allowed them to differentiate between projects and between criteria for the same project, in terms of the degree to which the criteria were met. The respondents suggested many specific changes to improve the description of a number of operationalisations, as well as adjustments to parts of the scoring form. They also gave valuable suggestions for a more precise procedure to apply Preffi, such as the suggestion to assess projects not only on the basis of the project description documents but to always include a consultation with the project manager to obtain information missing from the documentation and check interpretations of assessments. The study also showed that it was not always easy for the practitioners to apply Preffi 2.0, which means that users should preferably be trained in its use. It further showed that if several assessors judge the same project, they need to get together and discuss aspects for which their original assessments differed. The conclusions and suggestions of this study were incorporated in the definitive version of Preffi 2.0 and its user manual.

The findings of this first study did not allow hard conclusions to be drawn on the reliability of the draft version, as there was too little variance between the projects to enable us to determine generalisability coefficients (G coefficients) for many of the Preffi items.

A second study (chapter 6) tested the reliability of the definitive version of Preffi 2.0, in such a way as to avoid the limitations of the first study. It compared Preffi-based assessments of 20 projects, made by three practitioners, with intuitive assessments of the same projects they had given earlier and with intuitive assessments by three experts, which were used as an external criterion. To our surprise, the assessments by the three experts differed to such an extent that they could not really be used as an external criterion. The Preffi-based assessments showed acceptable reliability levels. On the basis of the scores for the individual criteria, two assessors would be necessary for a sufficiently reliable and accurate assessment of the project as a whole, while three assessors would be needed for a reliable and accurate assessment of each individual cluster. The overall scores per cluster were less reliable and accurate than would be acceptable.

This empirical study showed that a carefully planned development process aimed at developing a valid instrument does not necessarily result in a reliable

instrument. It showed that assessing health promotion projects is a complex process involving a great many aspects. Different assessors use different perspectives and include different aspects in their assessments. Since this is inevitable, a critical dialogue between them is useful and necessary to arrive at a balanced assessment of a project.

We found that generalisability theory offered a useful statistical technique to assess reliability and accuracy. The method estimates the number of assessors needed for a sufficiently reliable and accurate assessment, and also allows improvements to the instrument's reliability to be measured. For instance, the reliability and accuracy of the Preffi scores turned out to be better for the definitive version than for the draft version, thanks to the improvements made on the basis of the study of the draft version. The average number of assessors needed for sufficiently accurate assessments at the level of individual criteria fell from 8.13 for the draft version to 6.36 for the definitive version, while that at cluster level fell from 4 to 3. These two studies taught us that an instrument's reliability has to be won in a process of constant improvements, thus falsifying our naïve view at the start of the study that an instrument is immediately either reliable or unreliable. Empirical data challenge one to try and come up with improvements. We intend to go on improving the instrument, and regular follow-up research will have to show the effects of such improvements.

The studies have shown that discussions among assessors are needed to do justice to the various perspectives and produce a balanced judgement on a project. At a consensus meeting in our second study, the expert who had the most deviant assessments, thereby contributing greatly to the low reliability and accuracy of the experts' scores, regularly managed to win the other assessors over to his point of view. Health promotion involves many forms of expertise, relating to development, evaluation, implementation, etc. Our study shows that health promotion can benefit from this multidisciplinary approach, with contributions by many different forms of expertise. This also shows that there are various relevant forms of 'evidence', and that the notion of 'the' expert as a general, unspecified concept is not very useful and cannot be simply generalised, as is often assumed in statistical calculations on reliability and generalisability. When assessing the quality of complex projects like those in health promotion, it may therefore be worthwhile to require not only reliability estimations but also a critical dialogue between different assessors working from different perspectives.

The new Preffi version, with its operationalisations and norms, could be an attractive instrument for institutional managers and funding bodies to assess

projects and decide on their continuation. Our studies showed that such decision processes using Preffi 2.0 require a careful procedure, and we hope that the studies also contribute to the use of such careful procedures. This is important to avoid frustrating the implementation of the use of Preffi as a quality improvement instrument by health promotion specialists.

Our studies do leave certain questions unanswered. We found, for instance that three types of Preffi criteria (effect predictors) can be distinguished: criteria that allow an unambiguous assessment (e.g. 'Do the objectives specify the factors to be changed?'), criteria requiring judgements on aspects of the project (e.g. regarding the project manager's expertise) and criteria requiring an expert opinion (e.g., whether the theoretical model that was chosen is suitable). We did not find any significant differences in reliability between these three types of criteria, even though we would expect the second and third types to have lower reliability scores. The distinction between these three types had not been defined in advance as an explicit topic of research, and future studies should examine this distinction more specifically.

Another aspect that requires further research is the value of giving overall scores from 1 to 10 for each cluster as a whole and for the project as a whole. Our studies showed that the accuracy of these overall scores did not meet our predefined standard of $SEM < .26$. It would seem better to derive cluster scores strictly from the scores for the individual criteria. So far, Preffi offers the opportunity to deviate from this by giving greater weight to one or more of the criteria scores in calculating the cluster score. Our studies showed that this does not contribute to the reliability of the cluster scores. The first study found that there was no weighting effect to be observed at group level involving multiple assessors. This would seem to argue for deriving the overall cluster scores directly from the various criterion scores within each cluster. On the other hand, there would also seem to be valid arguments for allowing different weights. If, for instance, a new project does not involve a pre-test (6.7), this would have to result in a lower overall score for cluster 6 (intervention development) than would be produced by the arithmetic average.

Even though we have emphasised several times that Preffi's primary function is self-assessment for the purpose of improving projects, the above studies focused on the use of Preffi for assessment. Both studies, that of the draft version as well as that of the definitive version, found that the requirements for a diagnostic instrument might be less strict than those for an assessment instrument. Nevertheless, an instrument used for self-assessment will still need to be sufficiently reliable and accurate.

The entire Preffi project is of course based on the assumption that the use of the instrument will lead to actual improvements. We do indeed see such improvements in our capacity as consultants for the assessment of entries for the Preffi award, where Preffi criteria scores tend to be higher for the definitive entries. Similarly, the results of our Preffi training sessions indicate improvements. So far, however, there has not been any empirical research into this aspect. It would be useful to undertake a systematic study comparing Preffi-based assessments of a large number of projects prior to a programme of consultation or training with Preffi-based assessments after such a programme. Such a comparison could show which Preffi criteria refer to aspects that are easy or difficult to improve and what methods can be used to achieve such improvements.

7.5 Overall conclusion on Preffi 2.0

Preffi 2.0 is a valid instrument using evidence-based principles and guidelines for effect management. The second version is a useful further development of the first version and is useful and feasible for practitioners. The content validity of Preffi 2.0 is widely accepted and our studies have provided suggestions for sound ways to use Preffi for reliable assessments.

Preffi helps to diagnose the strong and weak points in projects and programmes. When applied to a number of projects or programmes together, it can even produce a rough diagnosis of the strong and weak aspects of all projects in a particular sector (see figure 1, chapter 5). This could allow specific improvement programmes to be developed. An empirical assessment of the question whether this would actually lead to better projects was beyond the scope of this PhD project.

Preffi also seems able to contribute to the purpose of synthesising new research findings and practical knowledge and making these accessible to practitioners.

It is our expectation that the general framework provided by Preffi can also contribute to neighbouring fields. One sector where the need for a general effectiveness framework is being actively discussed is that of youth care, where opinions are still divided about whether such a general framework could be found and what it should look like (Roede and Schooten, 2004).

7.6 Future prospects

The Preffi programme is part of a long-term strategy which involves producing regular updates of the instrument. We intend to introduce Preffi 3.0 in 2007. In developing this new version, we will once again attempt to improve the content and format of the instrument, to increase its reliability and to devise steps that might help integrate the consistent use of Preffi in existing operational processes.

7.6.1 Content

In developing the next Preffi version, the team will have to review the evidence supporting the various Preffi criteria, identify aspects for which new evidence has become available and decide how this can be used to further improve the content validity of the instrument. Literature research and suggestions made by colleagues will have to be used to decide what new effect predictors need to be incorporated in Preffi 3.0. Possible examples include greater attention to gender-specificity as a part of fitting interventions to target groups (6.3) (Voorham, 2003), elements of the capacity-building framework (Hawe et al., 1997; Hawe et al., 2000; Van den Broucke, 2003) which may provide a valuable supplement to the project management and capacity items, and Stacey's agreement certainty matrix, which may help to specify the commitment item (Stacey, 1999; Plsek and Greenhalgh, 2001; Kelly et al., 2004).

Some problematic topics and content aspects identified while using Preffi 2.0 will need to be further specified or solved. First, it is still a problem to apply scores to multiple interventions in one programme. A digital version of Preffi might make it technically easier to give multiple scores per programme. The question whether a programme includes the right mix of interventions, that is, whether there is sufficient coherence in the interventions (6.6), will have to be addressed in a further version of the instrument. It remains to be seen whether this will be possible within Preffi's general abstract frameworks, or whether it needs to be specified per topic or subject.

Second, some users have suggested giving the topic of recruitment a more prominent position in Preffi. In the current version, this is an element of implementation and project management.

Third, another interesting question raised by users is how Preffi should address the creativity factor. In the professionals' perception, programme or intervention characteristics such as 'fun', 'exciting' and 'innovative', in terms of concept

or format, contribute to the ultimate effect. Researchers have also pointed out that other aspects besides a systematic approach, such as habits and emotions, can play a role (Vries, 2000). Whether the creativity factor should become a new criterion is hard to say at this point in time. The Preffi team regards these as valuable elements, but wonders whether they can be specified on their own or are determined by a specific combination of existing Preffi criteria, such as timing, fitting in with the culture of the target group, using innovative methods, participation and simplicity. In any case, it is a topic that deserves to be addressed.

Capacity limitations have prevented us from including the topic of cost-effectiveness in the development of Preffi 2.0. The attention paid to monitoring and evaluation, which has been incorporated in Preffi 2.0, and which is beginning to take shape in practice as well, will have to be supplemented with the cost-effectiveness aspect in a later version. For one thing, this is being demanded by present-day society, which increasingly requires cost-effectiveness issues to be included in accountability. In addition, as the number of available and partly competing interventions grows, scarce resources for health promotion and disease prevention will make it increasingly urgent to assess how the smallest possible investment can yield the greatest possible effect among the largest possible percentage of the target population.

7.6.2 Format

One of Preffi's greatest assets is the apparently simple format of the instrument. However, users applying Preffi often want to explore certain criteria in greater depth and detail. Whereas Preffi's current format, on two sheets of A3-size paper, is not very suitable for this, a new, digital version of the instrument on the Internet might facilitate this. When a particular score is entered, the instrument might suggest tailored options for improvement, show examples of projects in which this particular aspect has been suitably addressed (modelling), or even refer to model programmes that relate specifically to this topic.

A digital version might also allow certain key concepts in health promotion to be linked more closely to Preffi. A concept like participation could be operationalised in the operational process used by health promotion specialists by being included in 17 criteria of Preffi 2.0, and the same could be done for concepts like empowerment and monitoring.

Ultimately, all Preffi criteria need to be applied to every project, but the order in which this is done may vary. A development programme has to meet differ-

ent requirements than an implementation programme. And a community approach requires greater emphasis on the target group's commitment at the start of a project than a general lifestyle campaign, which should focus on determinants of the behaviour to be influenced and on the local infrastructure that can be utilised. A digital version could facilitate this process for various types of projects, since such a version would offer more opportunities for tailoring interventions to the specific needs of an individual programme developer or a specific implementation context.

The desirability of a digital version has frequently been raised ever since Preffi 1.0 was launched in 1995. Some preliminary studies have been done, resulting in a project proposal that was favourably received in terms of quality but has not yet been implemented because of lack of funding. It seems inevitable, however, that further versions of Preffi will come in digital form.

7.6.3 Reliability

Preffi's reliability could be further improved by learning from users' experiences and accordingly changing the way it is used. This is a continuous process of learning. In the past few years, this process has involved applying Preffi as an evaluative framework in the Qui in-depth study on underprivileged groups in society (Poel, 2003). Preffi has further been used to select and work out model projects in the context of developing local intervention portfolios for specific topics that could support local health promotion specialists with certified interventions. Attempts are continuing to achieve external validation for Preffi. One of these attempts involves comparing jury assessments of projects submitted for the 2005 Preffi award with Preffi-based assessments of the same projects made by the Preffi team. These comparisons showed that the assessments corresponded very well.

The Preffi team expects that a digital version would contribute substantially to Preffi's reliability, as it could offer more specific explanations for each criterion, as well as suggestions that might lead to more objective assessments. It might provide links between those criteria that require expert judgement and expert systems like intervention mapping or databases with reviews and model programmes.

The basic concept behind Preffi is the generally accepted planning model, which was used to derive principles of effect management, which in turn were translated into criteria operationalised in general terms. Criteria formulated in

general terms can usually be interpreted in various ways, which clashes with the desire to achieve the most reliable assessment of those criteria. This problem can be partially solved by greater specification and better operationalisation of general effect predictors on the basis of further research, which will ultimately increase the reliability.

Another strategy to improve reliability, which is currently also being implemented, is to specify and apply the general principles included in Preffi for specific fields. One organisational consequence of this strategy was the merger between the Centre for Review and Implementation (CRIM) and the Preffi team to form the NIGZ Centre for Knowledge and Quality Management. The team is now collaborating with other NIGZ departments in developing new products provisionally known as health promotion packages and so-called local intervention portfolios for specific fields like obesity, smoking, etc. These products synthesise the information available from researchers, policymakers and practitioners, using Preffi as the structuring principle, and select examples of the most effective programmes and interventions in a particular field that could be eligible for certification. These packages are then offered to local health promotion specialists, accompanied by arguments supporting their content and a clear description of the contextual conditions that have to be met to allow the interventions to be implemented locally with the greatest possible effect. Health promotion specialists may compare these packages with their own local projects, which would allow them to evaluate their own practice in far more concrete terms. This should also contribute to the reliability of the Preffi-based assessments of their own projects.

7.6.4 Role and status

The implementation of Preffi has our continued attention, and its progress will be monitored regularly. This process should provide us with new ideas for further implementation, as well as indicating where Preffi is or is not being used. Thus, the so-called Trendonderzoek 2004 survey showed that, 15 months after Preffi 2.0 had been introduced, the instrument was being used in some way by 57% of Dutch health promotion specialists (Poel et al., 2005), particularly in national health promotion agencies, mental health care institutes (GGZ) and municipal health services (GGD).

The Preffi team intends to continue its implementation activities, in the form of publications, lectures, workshops, training courses and consultations. In addition to these activities, which aim at individual professionals, the best way

to strengthen the implementation process is by giving Preffi a permanent place in the day-to-day routine of health promotion specialists. Whether this will happen depends particularly on the commitment by the managers of the institutions that employ them, to whom this offers concrete opportunities to improve the quality of the work done at their institutions. The use of Preffi should be a part of this. The Preffi team has already developed a number of initiatives to achieve this in the past year, and these will be continued.

Specifically, these initiatives involve a joint programme of supplementary courses and continuing education that ties in with the NIGZ-NSPOH's basic package on health promotion, and includes courses on the various aspects of Preffi. The Foundation for the Harmonisation of External Quality Review in Health Care (HKZ) intends to draw up its certification scheme for health promotion next year. It is important that Preffi and the principles it incorporates are included in this scheme as the standard for the field. In addition, Preffi will be given an important place in the knowledge systems that are being developed or coordinated by the NIGZ Centre for Knowledge and Quality Management. Although this may not always be immediately evident, Preffi and the insights incorporated in it are part of the basic principles underlying the structure of and the questions addressed by the Qui project database, as well as the development of a portal for health promotion.

Providers of subsidies for health promotion projects, like the Dutch Ministry of Health, Welfare and Sports, municipal authorities and health insurers, want to know which interventions they should or should not fund. To this end, they would like to see a system of certification. NIGZ is currently drawing up requirements for the content of the protocols for certified interventions, partly on the basis of Preffi. Finally, Preffi is also to be incorporated in the implementation activities for the Referentiekader Gezondheidsbevordering (reference framework for health promotion) (Saan and de Haes, 2005)

The European Union is increasingly beginning to influence various activities in its member states, and this is also going to be the case for health promotion. It is expected that the recently founded European Centre for Disease Control and Prevention (ECDC) will have its own Health Promotion Department within 10 years, like its US equivalent CDC. It is within this context that the national public health and health promotion organisations in the EU countries have set up a long-term project called Getting Evidence into Practice, which is intended to result in a European action programme involving joint reviews and the development and implementation of European standards for the quality and effectiveness of health promotion. This project ties in with the Global Health

Promotion Effectiveness Program of the International Union for Health Promotion and Education (IUHPE). The European project is partly funded by the EU's Public Health programme, and is being coordinated by the NIGZ Centre for Knowledge and Quality Management. The project aims to develop a European Quality Assessment Tool (EQAT), to which we are making considerable contributions based on our experience with Preffi. The long-term prospect is that EQAT should become the European standard, and that the fourth or fifth version of Preffi should become part of, or become the Dutch version of, this European standard. For now, however, we are concentrating on the development of Preffi 3.0, for which we are learning valuable lessons from the international project.

The further strengthening of the role and status of Preffi also requires further research. The first topic of research should be the effects of the use of Preffi as a diagnostic instrument, in terms of the actual improvements it leads to. A second topic should be Preffi's predictive validity, that is, whether projects achieving high overall Preffi scores actually produce interventions that are demonstrably effective, or more effective than others. ZonMw has repeatedly indicated that it would like to see such a study undertaken, and the time seems right to actually do so.

Finally, we should not lose sight of Preffi's original perspective. Preffi is an instrument, a means to an end. The ultimate goal is to improve the effectiveness of health promotion in the Netherlands. While Preffi can certainly contribute to this, it is more important that all those involved join forces to create an infrastructure in which the quality of health promotion is the responsibility of many people, rather than of researchers and health promotion specialists alone.

References

- Ader, M., Berensson, K., Carlsson, P., Granath, M. and Urwitz, V. (2001) Quality indicators for health promotion programmes *Health Promot. Int.*, 16, 187-195.
- Airhihenbuwa, C. O. (1995) *Health and culture: beyond the Western paradigm*. Sage, Thousand Oaks.
- Ajzen, I. (1988) *Attitudes, personality and behavior*. Open University Press, Milton Keynes, UK.
- Ajzen, I. (1991) The theory of planned behavior *Organizational Behavior and Human Decision Processes*, 50, 179-211.
- Assema, P. v., Brug, J., Glanz, K., Dolders, M. and Mudde, A. (1998) Nationwide implementation of guided supermarket tours in the Netherlands: a dissemination study *Health Educ. Res.*, 13 (4), 557-566.
- Australian Health Promotion Association (2001) Health Promotion and Evaluation ; a programmatic approach *Health promotion Journal of Australia*, 11, Supplement 1-28.
- Bandura, A. (1986) *Social foundations of thoughts and actions: a social cognitive theory*. Prentice-Hall, Englewood Cliffs.
- Barnhorn, H. and Walda, R. (1992) *De eerste lijn op spitzen: zorgvernieuwing in de praktijk*. Van Gorkum, Assen.
- Baron-Epel, O., Levin-Zamir, D., Satran-Argaman, C., Livny, N. and Amit, N. (2004) A participatory process for developing quality assurance tools for health education programs *Patient Education and Counseling*, 54, 213-219.
- Bartholomew, L. K., Parcel, G. S., Kok, G. and Gottlieb, N. H. (2001) *Intervention mapping: designing theory- and evidence-based health promotion programs*. Mayfield, Mountain View.
- Bekkering, T., Glas, H., Klaassen, D. and Walter, J (2001) *Management van processen, succesvol realiseren van complexe initiatieven*. Het Spectrum, Utrecht.
- Berensson, K., Granath, M. and Urwitz, V. (1996) *Succeeding with Health Promotion-Quality Assurance*. Landstingsförbundet, Stockholm.
- Boddy, D. (Ed.) (1999) *The evidence of health promotion effectiveness; shaping public health in New Europe*, IUHPE & European Commission, Paris-Luxembourg.
- Bos, J. and Harting, E. (1998) *Projectmatig creëren*. Scriptum Books, Schiedam.

- Bosma, M. W. M. and Hosman, C. M. H. (1990) *Preventie op waarde geschat. Een studie naar de beïnvloedbaarheid van determinanten van psychische gezondheid*. Beta Boeken, Nijmegen.
- Botvin, G. J. (2004) Advancing Prevention Science and Practice: Challenges, critical issues and future directions *Prevention Science*, 5, 69-72.
- Bracht, N. (1999) *Health promotion at the community level: new advances (2nd ed.)*. Sage, Thousand Oaks.
- Brown, C. H., Berndt, D., Brinales, J., Zong, X. and Bhagwat, D. (2000) Evaluating the evidence of effectiveness for preventive interventions: using a registry system to influence policy through science *Addictive Behaviors*, 25, 955-964.
- Brug, J., Schaalma, H., Kok, G., Meertens, R. M. and Molen, H. T. v. d. (2000) *Gezondheidsvoorlichting en gedragsverandering, een planmatige aanpak*. Van Gorcum, Assen/Maastricht.
- Burgers, J., Grol, R. and Zaat, J. (2001) Richtlijnen en protocollen als hulpmiddel bij de verbetering van de zorg. In Grol, R. and Wensing, M. (eds), *Implementatie; Effectieve verandering in de patientenzorg*. Elsevier gezondheidszorg, Maarssen.
- Burgers, J. S., Grol, R., Klazinga, N., Mäkelä, M. and Zaat, J. (2003) Towards evidence-based clinical practice: an international survey of 18 clinical guideline programs *International Journal for Quality in Health Care*, 15, 31-45.
- Butterfoss, F. D., Francisco, V. T. and Capwell, E. M. (2001) Stakeholder participation in evaluation *Health Promotion Practice*, 2, 114-119.
- Cameron, R., Jolin, M. A., Walker, R., McDermott, N. and Gough, M. (2001) Linking Science and practice: Toward a system for enabling communities to adopt best practices for chronic disease prevention *Health Promotion Practice*, 2, 35-42.
- Christiansen, G. (1999) *Evaluation, Ein Instrument zur Qualitätssicherung in der Gesundheitsförderung*. Bundeszentrale für gesundheitliche Aufklärung (BzGA), Köln.
- CMHA and Health Canada (1999) *Canadian Mental Health Promotion toolkit*.
- Cochrane Collaboration. www.cochrane.org, website of the Cochrane organisation on systematic reviews.
- Commissie Dunning (1991) *Kiezen en Delen, rapport van de commissie Keuzen in de zorg (Commissie Dunning)*. Ministerie van WVC, Den Haag.
- CSAP (Center for Substance Abuse Prevention) (2002) *Achieving outcomes; a practitioner's guide to effective prevention*. CSAP/SAMHSA, Rockville MD.

- CSAP (Center for Substance Abuse Prevention) Prevention Enhancement Protocols System (PEPS) official document, sub-website of preventionondss.com providing information on the review methodology.
- Cuijpers, P. (2000) Effective ingredients of school-based drug prevention programs. A systematic review. *Addictive Behaviors*, 27, 1009-1023.
- Damoiseaux, V., Molen, H. T. v.d. and Kok, G. J. (Eds.) (1993) *Gezondheidsvoorlichting en gedragsverandering*, van Gorcum/Open Universiteit, Assen/Heerlen.
- Davies, J. K. and Macdonald, G. (Eds.) (1998) *Quality, Evidence and Effectiveness in health promotion*, Routledge, London.
- Dijkster, A., Dongen, M. V. and Brug, J. (2000) *Gezondheidsvoorlichting en gedragsverandering: een planmatige aanpak*. Van Gorcum, Assen.
- Donabedian, A. (1980) Explorations in quality assessment and monitoring. Vol 1: The definitions of quality and approaches to its assessment.
- Driel, W. G. v. and Keijsers, J. F. E. M. (1997) An instrument for reviewing the effectiveness of health education and health promotion *Patient Education and Counseling*, 30, 7-17.
- Driel, W. G. v., Keijsers, J. F. E. M., Zaal, K. and Güttinger, V. (1994) *Effectiviteitsanalyse van gvo/ggz-preventie interventies, Ontwikkeling van een analyse-instrument*. Landelijk Centrum GVO, Utrecht.
- Durlak, J. A. (1998) Why program implementation is important *Journal of Prevention and Intervention in the Community*, 17, 5-18.
- Durlak, J. A. and Welsh, A. M. (1997) Primary prevention mental health programs for children and adolescents: A meta-analytic review *Am J Commun Psychology*, 25, 115-152.
- Dusenbury, L. and Falco, M. (1995) Eleven components of effective drug abuse prevention curricula *Journal of School Health*, 65, 420-425.
- Eggen, J. H. M. and Sanders, P. F. (Eds.) (1993) *Psychometrie in de praktijk*, Citogroep, Arnhem.
- Evans, D., Head, M. J. and Speller, V. (1994) *Assuring Quality in Health Promotion. How to Develop Standards of Good Practice*. Health Education Authority, London.
- Fleuren, M., Wiefferink, K. and Paulussen, T. (2004) Determinants of innovation within health care organizations: Literature review and Delphi study *Int J Qual Health Care*, 16, 107-123.
- Francisco, V. T., Capwell, E. M. and Butterfoss, F. (2000) Evaluation in practice: getting off to a good start with your evaluation *Health Promotion Practice*, 1, 126-131.

- French, J. (2004) *The learning from Effective Practice Standard System (LEPSS)*. Health Development Agency, London.
- Gageldonk, A. v. and Cuijpers, P. (1998) *Effecten van verslavingspreventie: overzicht van de literatuur*. Trimbo-instituut, Utrecht.
- Gevers, J. K. M. and Biesart, M. C. I. H. (1999) Medische beslissingen, kostenoverwegingen en richtlijnen voor medisch handelen: kanttekeningen vanuit het recht *Nederlands tijdschrift voor Geneeskunde*, 143, 2629-2632.
- Glanz, K., Lewis, F. M. and Rimer, B. K. (Eds.) (1997) *Health behavior and health education: theory, research and practice*. Jossey-Bass, San Francisco.
- Glanz, K., Lewis, F. M. and Rimer, B. K. (Eds.) (2002) *Health behavior and health education: theory, research and practice*, Jossey-Bass, San Francisco.
- Godin, G. and Kok, G. (1996) The theory of planned behavior: a review of its applications to health-related problems *American Journal of Health Promotion*, 11, 87-98.
- Green, L. W. and Kreuter, M. W. (1991) *Health promotion planning, an educational and environmental approach*. Mayfield, Mountain View.
- Green, L. W. and Kreuter, M. W. (1999) *Health promotion and planning: an educational and ecological approach*. Mayfield, Mountain View.
- Green, L. W. and Kreuter, M. W. (2004) *Health Program Planning: an educational and ecological approach*. McGraw-Hill.
- Greenberg, M. T., Domitrovich, C. E., Graczyk, P. and Zins, J. (2001) *A Conceptual Model of Implementation for School-Based Preventive Interventions: Implications for Research, Practice, and Policy*. Center for Mental Health Services, SAMHSA, Rockville.
- Grimshaw, J. M., Thomas, G. E., MacLennan, G., Fraser, C., Ramsay, C. R., Vale, L., Whitty, P., Eccles, M. P., Matowe, L., Shirran, L., Wensing, M., Dijkstra, R. and Donaldson, C. (2004) Effectiveness and efficiency of guideline dissemination and implementation strategies *Health Technology Assessment*, 8, 1-72.
- Grol, R. T. P. M. (1994) *Invoering van richtlijnen en veranderingen*. De Tijdstroom, Utrecht.
- Grol, R. T. P. M. and Wensing, M. (Eds.) (2001) *Implementatie, effectieve verandering in de patiëntenzorg*. Elsevier Gezondheidszorg, Maarssen.
- Groote, G. P. (1997) *Projecten leiden, methoden en technieken voor projectmatig werken*. Het Spectrum/Marka, Utrecht.
- Haglund, B. J. A., Jansson, B., Petterson, B. and Tillgren, P. (1998) A quality assurance instrument for practitioners. An example from Sweden. In Davies, J. K. and Macdonald, G. (eds), *Quality, Evidence and Effectiveness in health promotion*. Routledge, London, p. 93-116.

- Hawe, P., Noort, M., King, L. and Jordens, C. (1997) Multiplying health gains: the critical role of capacity building within health promotion programs *Health Policy*, 29-42.
- Hawe, P., King, L., Noort, M., Jordens, C. and Lloyd, B. (2000) *Indicators to help with capacity building in health promotion*. New South Wales Health Department, Sydney.
- Health Canada (2001) *The Population Health Template: Key Elements and Actions That Define a Population Health Approach*.
- Health Switzerland (2002) www.quint-essenz.ch/, Quint-essenz 4. 0, a website supporting the design of health promotion projects.
- Hendricks, K. B. and Singhal, V. R. (1997) Does implementing an effective TQM program actually improve operating performance? Empirical evidence from firms that have won Quality Awards. *Management Science*, 43, 1258-1274.
- Heuvelmans, A. P. J. M. and Sanders, P. F. (1993) Beoordelingsovereenstemming. In Eggen, J. H. M. and Sanders, P. F. (eds), *Psychometrie in de praktijk*. Citogroep, Arnhem, p. 443-473.
- Hommels, L. M., Molleman, G. R. M. and Kok, H. H. (2000) Handboek Preffi. NIGZ, Woerden.
- Hosman, C. M. H., van Doorn, I. and Verburg, H. (Eds.) (1988) *Preventie in-zicht.*, Swets en Zeitlinger, Lisse.
- Hosman, C. M. H. (1995) Effectiveness and effect management in mental health promotion and prevention. In Trent, D. R. and Reeds, C. A. (eds), *Promotion of mental health*. Avebury, Aldershot, p. 5-23.
- Hosman, C. M. H. and Engels, M. C. L. J. (1999) The value of model programmes in Mental Health Promotion and Mental Disorder Prevention *Int J Mental Health Prom.*, 4-16.
- Hosman, C. M. H. and Llopis, E. J. (1999) Evidence of mental health promotion effectiveness. In Boddy, D. (ed), *The evidence of health promotion effectiveness: shaping public health in New Europe*. IUHPE & European Commission, Paris- Luxembourg, p. 29-41.
- Hosman, C. M. H. (2000) Quality criteria for model programmes. In Laivikainen, J., Lahtinen, E. and Lehtinen, V. (eds), *Public health approach on mental health in Europe*. Stakes, Helsinki, p. 109-115.
- Hosman, C. M. H. and Clayton, R. (2000) Prevention and health promotion on the international scene: The need for a more effective and comprehensive approach *Addictive Behaviors*, 25, 943-954.
- Hosman, C. M. H. and Llopis, E. (2002) *A conceptual framework for predicting and enhancing effectiveness of prevention and health promotion; the case of mental health*. Prevention Research Centre, Nijmegen.

- Hosman, C. M. H., Llopis, E. J. and Saxena, S. (2004) *Prevention of Mental disorders : Effective interventions and policy options*. World Health Organisation, Geneva.
- Huff, R. M. and Kline, M. V. (1999) *Promoting health in multicultural populations: a handbook for practitioners*. Sage, Thousand Oaks.
- Hulscher, M., Wensing, M. and Grol, R. (2000) *Effectieve implementatie: theorieën en strategieën*. ZorgOnderzoek Nederland, Den Haag.
- IOM (2001) *Crossing the Quality Chasm: The IOM Health Care Quality initiative*. Institute of Medicine, Washington DC.
- Jansen, J., Schuit, A. and Lucht, F. v. d. (Eds.) (2002) *Tijd voor gezond gedrag. Bevordering van gezond gedrag bij specifieke groepen.*, RIVM/Bohn Stafleu Van Loghum, Bilthoven/Houten.
- Janssen, M. and Geelen, K. (1994) *Beter Voorkomen, kwaliteitscriteria voor preventieprojecten in de geestelijke gezondheidszorg*. Nationaal centrum Geestelijke volksgezondheid, Utrecht.
- Jones, S. C. and Donovan, R. J. (2004) Does theory inform practice in health promotion in Australia? *Health Educ. Res.*, 19, 1-14.
- Jong, M. D. T. de and Schellens, P. J. (2000) Formatieve evaluatie. In Schellens, P. J., Klaassen, R. and Vries, S. d. (eds), *Communicatieontwerpen. Methoden, perspectieven en toepassingen*. Van Gorcum, Assen.
- Jonkers, R., Liedekerken, P. C., Haes, W. F. M. d., Kok, G., J. and Saan, J. A. M. (Eds.) (1988) *Effectiviteit van Gezondheidsvoorlichting en -opvoeding (GVO)*, Uitgeverij voor Gezondheidsbevordering, Rijswijk.
- Kahan, B. and Goodstadt, M. (2001) The Interactive Domain Model of Best Practices in Health Promotion: Developing and Implementing a Best Practices Approach to Health Promotion *Health Promot Pract*, 2, 43-67.
- Keijsers, J. F. E. M. and Saan, J. A. M. (1998) The development of two instruments to measure the quality of health promotion interventions. In Davies, J. and Macdonald, G. (eds), *Quality, Evidence and Effectiveness in health promotion*. Routledge, London, p. 117-129.
- Keijsers, J. F. E. M. and Vaandrager, L. (Eds.) (2000) *Gezond leven. Stand van zaken en voorstel voor programmering*, ZorgOnderzoek Nederland, Den Haag.
- Kelly, M. P., Speller, V. and Meyrick, J. (2004) Getting evidence into practice in public health. Health Development Agency, London.
- King, L., Hawe, P. and Wise, M. (1998) Making dissemination a two-way process *Health Promotion International*, 13, 237-244.

- Kliche, J., Töppich, J., Kawski, S., Koch, U. and Lehmann, H. (2004) Die Beurteilung der Struktur-, Konzept- und Prozessqualität von Prävention und Gesundheitsförderung; Anforderungen und Lösungen *Bundesgesundheitsbl-Gesundheitsforsch - Gesundheitsschutz*, 47, 125-132.
- Koelen, M. A., Vaandrager, L. and Colomèr, C. (2001) Health promotion research: dilemmas and challenges *Journal of Epidemiology and Community Health*, 55, 257-262.
- Koelen, M. A. and Ban, A. W. v. d. (2004) *Health Education and Health Promotion*. Wageningen Academic Publishers, Wageningen.
- Kok, G. J. (1985) Een model van gedragsverandering via voorlichting *Nederlands tijdschrift voor psychologie*, 40, 71-76.
- Kok, G. (1992) Quality of planning as a decisive determinant of health education effectiveness *Hygie*, 11, 3-9.
- Kok, G. J. and Borne, H. W. v. d. (1995) Effectiviteit van gezondheidsvoorlichting en gezondheidsbevordering; Meta-analyses van effect-studies en determinanten van effectiviteit *TSG, tijdschrift voor gezondheidswetenschappen*, 73, 347-353.
- Kok, G. J., Borne, B. v. d. and Mullen, P. D. (1997) Effectiveness of health education and health promotion: meta-analyses of effect studies and determinants of effectiveness. *Patient Education and Counseling*, 30, 19-27.
- Kok, G. (1999) Implementing Mental Health Promotion: a health education and promotion perspective *Int J Mental Health Prom.*, 1, 4-9.
- Kotler, P. and Roberto, E. (1991) *Sociale marketing*. Het Spectrum, Utrecht.
- Laverack, G. and Labonte, R. (2000) A planning framework for community empowerment goals within health promotion *Health Policy and Planning*, 15, 255-262.
- Llopis, E. (2002) *What makes the ounce of prevention effective? A meta-analysis of mental health promotion and mental disorder prevention*. Quickprint, Nijmegen.
- Llopis, E. J., Hosman, C. M. H., Kasander, R. and Walhbeck, K. (2004) *International database of mental health promotion and mental disorder prevention programmes and policies: IMHPA Database*. Prevention Research Centre, University of Nijmegen, Nijmegen.
- Llopis, E. J., Hosman, C. M. H. and Brown, C. H. (submitted) Gender and age as population predictors of effect in a meta-analysis of mental disease prevention and mental health promotion programs across the life span. (submitted).
- Lomas, J. (1997) *Improving research dissemination and uptake in the health sector: beyond the sound of one hand clapping*. Centre for Health Economics and Policy Analysis, McMaster University, Hamilton, Canada.

- Macdonald, G. (1996) *Presentation of existing standards for evaluation and quality assessment of health promotion intervention in the 15 EU Member States*. IUHPE/Euro & NIGZ, Woerden.
- McGuire, W. J. (1985) Attitudes and attitude change. In Lindsay, G. and Aronson, E. (eds), *The handbook of social psychology*. Random House, New York, p. 233-346.
- McKay, A. (2000) Prevention of sexually transmitted infections in different populations: a review of behaviourally effective and cost-effective interventions. *The Canadian Journal of Human Sexuality*, 9, 95-120.
- McQueen, D. V. (2000) Perspectives on health promotion: theory, evidence, practice and the emergence of complexity *Health Promotion International*, 15, 95-97.
- McQueen, D. V. and Anderson, L. M. (2002) What counts as evidence: issues and debates on evidence relevant to the evaluation of community health programs. In Rootman, I., Goodstadt, M., McQueen, D., Potvin, L., Springett, J. and Ziglio, E. (eds), *Evaluation in Health Promotion: principles and perspectives*. WHO/EURO, Copenhagen.
- Meertens, R., Schaalma, H., Brug, J. and Vries, N. de. (2000) Determinanten van gedrag. In Brug, J., Schaalma, H., Kok, G., Meertens, R. M. and Molen, H. T. v. d. (eds), *Gezondheidsvoorlichting en gedragsverandering, een planmatige aanpak*. Van Gorkum, Assen, p. 55-72.
- Mesters, I., Kok, G. J. and Schaalma, H. (2000) Diffusie van gezondheidsvoorlichtingsprogramma's. In Brug, J., Schaalma, H., Kok, G., Meertens, R. M. and Molen, H. T. v. d. (eds), *Gezondheidsvoorlichting en gedragsverandering, een planmatige aanpak*. Van Gorkum, Assen, p. 119-132.
- Meurs, L. H. v. (2002) *Concept Preffi 2.0; reliability and usefulness, research paper (in Dutch)*. NIGZ, Woerden.
- Michelbrink, F. (2000) *Praktijkegericht onderzoek in zorg en welzijn*. SWP, Utrecht.
- Ministerie van Financiën (2002) *Van beleidsbegroting tot beleidsverantwoording*. Min. van Financiën, Den Haag.
- Ministerie van Welzijn Volksgezondheid en Cultuur (1992) *Preventiebeleid voor de volksgezondheid*. WVC, Rijswijk.
- Minkler, M. and Wallerstein, N. (1996) *Health behavior and health education*. Jossey-Bass, San Francisco.
- Molleman, G. R. M., Nies, H. A. M. v., Olthof, D. F. L. and Welling, B. J. M. (1992) *Trendonderzoek NVPG 1991*. NVPG, Amersfoort.
- Molleman, G. R. M. and Nies, H. A. M. v. (1995) *Trendonderzoek NVPG 1994*. NVPG, Nijmegen/Leiden.

- Molleman, G. R. M., Driel, W. v. and Keijzers, J. F. E. M. (1995) *Preventie Effectiviteits-instrument, PREFFI 1.0. Ontwikkeling van een effectiviteitsinstrument voor de gvo/preventiepraktijk*. Landelijk Centrum GVO, Utrecht.
- Molleman, G. R. M. (1999) Implementing the Preffi: the use of guidelines for practitioners in the Netherlands. In Norheim, L. and Waller, M. (eds), *Best Practices, a selection of papers on Quality and Effectiveness in Health Promotion*. Finnish Centre for Health Promotion, Helsinki/Talinn.
- Molleman, G. R. M., Kok, H. H. and Hommels, L. M. (2000) *Preffi, Brug tussen kennis en praktijk*. NIGZ, Woerden.
- Molleman, G. R. M., Peters, L. W. H., Hommels, L. M. and Ploeg, M. A. (2003) *Assessment Package; Health Promotion Effect Management Instrument Preffi 2.0*. NIGZ, Woerden, Netherlands.
- Molleman, G. R. M. and Hosman, C. M. H. (2003) Ontwikkeling van een kwaliteitsinstrument voor de effectiviteit van gvo/preventie-programma's; de Preffi 1.0, ontwikkeling en ervaringen en uitgangspunten voor een Preffi 2.0. *TSG Tijdschrift voor gezondheidswetenschappen*, 81, 238-246.
- Molleman, G. R. M., Peters, L. W. H., Hosman, C. M. H., Hommels, L. M., Ploeg, M. A., Kok, H. H. and Llopis, E. J. (2003) De Preffi 2.0: systematische ontwikkeling van een kwaliteitsinstrument voor gezondheidsbevoorde-ring *TSG Tijdschrift voor gezondheidswetenschappen*, 81, 247-255.
- Molleman, G. R. M., Peters, L. W. H., Hosman, C. M. H. and Kok, G. J. (2004) Implementation of a quality assurance instrument (Preffi 1.0) to improve the effectiveness of health promotion in The Netherlands *Health Educ. Res.*, cyg142.
- Molleman, G. R. M., Ploeg, M. A., Hosman, C. M. H. and Peters, L. W. H. (2004) Preffi 2.0 : un outil néerlandais pour analyser l'efficacité des interventions en promotion de santé *Promotion & Education*, 22-27.
- Montano, D. E. and Kasprzyk, D. (2002) The theory of reasoned action and the theory of planned behavior. In Glanz, K., Rimer, B. K. and Lewis, F. M. (eds), *Health Behavior and Health Education, 3rd edition*. Jossey-Bass, San Francisco CA, p. 67 - 98.
- Mrazek, P. and Hosman, C. M. H. (2002) *Towards a strategy for worldwide action to promote mental health and prevent mental and behavioral disorders*. World Federation for Mental Health, Carter Center, US Center for Mental Health Services, Atlanta.
- Mullen, P. D. (1985) Clinical trials of patient education for chronic conditions: a comparative meta-analysis of intervention types *Preventive Medicine*, 14, 753-781.

- Nabitz, U., Klazinga, N. and Walburg, J. (2000) The EFQM excellence model: European and Dutch experiences with the EFQM approach in health care *Int J Qual Health Care*, 12, 191-202.
- NHS Centre for Reviews and Dissemination (CRD) York University. www.york.ac.uk/inst/crd website of this British review centre.
- NIGZ (2003) www.preff.nl website NIGZ Centre for Knowledge and Quality Management, Woerden.
- NIZW (2005) www.jeugdinterventies.nl website van Databank Effectieve Jeugdinterventies. NIZW-jeugd, Utrecht
- Nunnally, J. C. (1967) *Psychometric theory*. McGrawHill., New York.
- Nutbeam, D. (1998) Evaluating health promotion: progress, problems and solutions *Health Promotion International*, 13, 27-44.
- Nutbeam, D. (1999) Evaluatie van oorzaak en gevolg in gezondheidsbevorderingsprojecten: aanpassing van onderzoeksmethoden aan interventiemethoden *TSG Tijdschrift voor gezondheidswetenschappen*, 77, 15-23.
- NVGVO (1988) *Effectiviteit van GVO en preventie, Wilhelmina Rouwenhorst Lezing*. NVGVO, Maastricht.
- NVPG (2001) *Beroepsprofiel preventie en gezondheidsbevordering 2000-2003*. NVPG, Woerden.
- Oldenburg, B., Hardscastle, D. M. and Kok, G. (1997) Diffusion of innovations. In Glanz, K., Lewis, F. M. and Rimer, B. K. (eds), *Health Behavior and Health Education: Theory, Research and Practice*. Jossey-Bass, San Francisco.
- Oldenburg, B. and Parcel, G. S. (2002) Diffusion of innovation. In Glanz, K., Rimer, B. K. and Lewis, F. M. (eds), *Health Behavior and Health Education, 3rd edition*. Jossey-Bass, San Francisco CA, p. 312-334.
- Orlandi, M. A., Landers, C., Raymond, W. and Haley, N. (1990) Diffusion of Health Promotion Innovations. In Glanz, K., Rimer, B. K. and Lewis, F. M. (eds), *Health Behavior and Health Education, Theory, Research and Practice*. Jossey-Bass, San Francisco CA, p. 288-314.
- Ovreteit, J. (1996) Quality in health promotion *Health Promot. Int.*, 11, 55-62.
- Oxman, A. D., Thomson, M. A., Davis, D. A. and Haynes, R. B. (1995) No magic bullets: a systematic review of 102 trials of interventions to improve professional practice *Canadian Medical Association Journal*, 153, 1423-1431.
- Parcel, G. S., Perry, C. L. and Taylor, W. C. (1990) Beyond demonstration; diffusion of health promotion innovations. In Bracht, N. (ed), *Health promotion at the community level*. SAGE, Newbury Park, CA.
- Paulussen, T. G. W. (1994) *Adoption and implementation of AIDS education in Dutch secondary schools*. Landelijk Centrum GVO, Utrecht.

- Paulussen, T., Aarts, H., Leeuw, E. de. and Kok, G. J. (1997) Werkzaamheid en doeltreffendheid van gezondheidsvoorlichting bij roken, lichamelijke activiteit en stresshantering. In Gunning-Schepers, L. J. and Jansen, J. (eds), *Volksgezondheid Toekomst Verkenningen 1997, Effecten van preventie, deel IV*. RIVM, Elsevier/De Tijdstroom, Bilthoven, Maarssen.
- Peters, L. W. H. and Keijsers, J. F. E. M. (2002) *Kwaliteitscriteria voor gezondheidsbevorderende interventies in en om scholen*. NIGZ & Schoolslag, Zuid Limburg, Woerden.
- Peters, L. W. H., Molleman, G. R. M. and Hielkema, J. E. (2000) *Preffi, implementatietraject 1997–1999; onderzoeksverslag*. NIGZ, Woerden.
- Peters, L. W. H., Molleman, G. R. M., Hommels, L. M., Ploeg, M. A., Hosman, C. M. H. and Llopis, E. (2003) *Explanatory Guide Preffi 2.0*. NIGZ, Woerden, Netherlands.
- Pieterse, M. (1999) *Stoppen met roken met hulp van de huisartspraktijk. Ontwikkeling, toetsing en implementatie van een minimale interventiestrategie*. Universiteit Twente, Enschede.
- Poel, F. v. d. (2003) *Wat laat het GVO/preventieveld ons zien? Overview of health promotion activities in the Netherlands*. NIGZ, Woerden.
- Poel, F. v. d., Waisvisz, F. and Molleman, G. R. M. (2005) *Trendonderzoek Gezondheidsbevordering en Preventie 2005*. NIGZ Centrum voor Kennis en Kwaliteit, Woerden.
- Pretty, J. N. (1995) *Regenerating agriculture: policies and practice for sustainability and self-reliance*. Earthscan Publications Ltd., London.
- Prochaska, J. O., DiClemente, C. C. and Norcross, J. C. (1992) In search of how people change: Applications to addictive behaviors *American Psychologist*, 1102-1114.
- Prochaska, J. O., Redding, C. A. and Evers, K. E. (2002) The transtheoretical model and stages of change. In Glanz, K., Rimer, B. K. and Lewis, F. M. (eds), *Health Behavior and Health Education, 3rd edition*. Jossey-Bass, San Francisco CA, p. 99-120.
- Public Health and Aged Community and Mental health (2000) *Primary Care Partnership; draft Health Promotion Guidelines*. Victorian Government Department of Human Services, Melbourne.
- Rada, J., Ratima, M. and Howden-Chapman, P. (1999) Evidence-based purchasing of health promotion: methodology for reviewing evidence *Health Promotion International*, 14, 177-187.
- Ridder, D. de. (2003) *Verre doelen, onmiddellijke frustraties: Zelfregulatieprocessen in gezondheidsgedrag*. R.U. Utrecht, Utrecht.

- Rifkin, S. B., Lewando-Hundt, G. and Draper, A. K. (2000) *Participatory approaches in health promotion and health planning: a literature review*. Health Development Agency, London.
- Riley, T. and Hawe, P. (2004) Researching practice: the methodological case for narrative inquiry *Health Educ. Res.*, cyg122.
- Roede, E. and Schooten, E. v. (2004) *Kiezen voor effectiviteit: haalbaarheid van een uniforme effectiviteitstoets voor preventieve, curatieve en repressieve jeugdprogramma's*. SCO-Kohnstamm Instituut, Amsterdam.
- Rogers, E. (1995) *Diffusion of innovations*. The Free Press, New York.
- Roling, N. G., Kuiper, D. and Janmaat, R. (Eds.) (1994) *Basisboek Voorlichtingskunde*, Boom, Amsterdam Meppel.
- Rootman, I., Goodstadt, M., McQueen, D., Potvin, L., Springett, J. and Ziglio, E. (2002) *Evaluation in Health Promotion: principles and perspectives*. WHO/EURO, Copenhagen.
- Saan, J. A. M. (1993) *Actieprogramma Kwaliteit*. Landelijk Centrum GVO, Utrecht.
- Saan, J. A. M. (1999) Assessment van sociale technologie in 'public health' *TSG Tijdschrift voor gezondheidswetenschappen*, 77, 5-14.
- Saan, J.A.M., Molleman, G. R. M., Vermeer, V. and Dale, D. v. (2004) *Projectmanagement: het opus-model*. NIGZ, Woerden.
- Saan, J. A. M. and Haes, W. F. M. de. (2005) *Gezond effect bevorderen*. NIGZ, Woerden.
- SAMSHA and US Department of Health and Human Services (2002) *SAMSHA Model Programs National Dissemination System*. SAMSHA, Rockville.
- Sanderson, C., Haglund, B. J. A., Tillgren, P., Svanström, L., Östenson, C. G., Holm, L. E., Ullen, H. and Smajkic, A. (1996) Effect of stage models in community intervention programmes; and the development of the Model for Management of Intervention Programma Preparation (MMIPP) *Health Promotion International*, 11, 143-156.
- Scheerder, G., Van den Broucke, S. and Saan, H. (2003) *Projecten voor gezondheidspromotie; een handleiding voor kwaliteitsvol werken*. Garant, Antwerpen - Apeldoorn.
- Schinke, S., Brounstein, P. and Gardner, S. (2002) *Science-Based Prevention Programs and Principles, 2002*. Center for Substance Abuse Prevention, SAMHSA, Rockville, MD.
- Shrout, P. E. and Fleiss, J. L. (1979) Intraclass correlations: Uses in assessing rater reliability *Psychological Bulletin*, 86, 420-428.
- Shavelson, R. J. and Webb, N. M. (1991) *Generalizability theory: a primer*. Sage Publications, Newbury Park, CA.

- Shonkoff, J. P. (2000) Science, policy, and practice: three cultures in search of a shared mission *Child Development*, 71, 181-187.
- South East Health (2003) *Four Steps Towards Equity*. Health Promotion Services, South East Health, Sidney.
- Speller, V., Evans, D. and Head, M. J. (1997) Developing quality assurance standards for Health Promotion practice in the UK *Health Promotion International*, 12 (3), 215-224.
- Speller, V., Rogers, L. and Rushmer, A. (1998) Quality assessments in health promotion settings. In Davies, J. K. and Macdonald, G. (eds), *Quality, Evidence and effectiveness in health promotion*. Routledge, London, p. 130-146.
- Splunteren, P. T. v. and Borselaar, H. (1998) *Werken aan vernieuwing: handreikingen voor ontwikkeling en implementatie van vernieuwingen in de ouderenzorg*. ZorgOnderzoek Nederland/Ipsos Facto, Den Haag.
- Splunteren, P. v. (1999) De kunst van het implementeren is het voeren van een goede dialoog *Mediator*, 1, 20-21.
- Stacey, R. D. (1999) *Strategic management and organisational dynamics: the challenge of complexity*. Financial Times/Prentice Hall, New York.
- Steckler, A., Goodman, R. M. and Kegler, M. C. (2002) Mobilizing organisation for health enhancement: theories of organisation change. In Glanz, K., Rimer, B. K. and Lewis, F. M. (eds), *Health Behavior and Health Education*, 3rd edition. Jossey-Bass, San Francisco CA, p. 335-360.
- Swanborn, P. G. (1999) *Evalueren*. Boom, Amsterdam.
- The Agree Collaboration (2001) Appraisal of guidelines for Research & Evaluation (AGREE) Instrument www.agreecollaboration.org.
- Theunissen, N. C. M., Pas, S. M. E. t. and Friele, R. F. (2003) *Evaluatie implementatie Monitor ZonMw*. NIVEL, Utrecht.
- Tobler, N. S. and Stratton, H. H. (1997) Effectiveness of school-based drug prevention programs: a meta-analysis of the research. *The Journal of Primary Prevention* 18, 1, 71-128.
- Tones, K. (1999) De evaluatie van gezondheidsbevordering: een verhaal van drie fouten. TSG *Tijdschrift voor gezondheidswetenschappen*, 77, 24-28.
- Tones, K. and Tilford, S. (1994) *Health education, effectiveness, efficiency and equity*. Chapman & Hall, London.
- US Department of Health and Human Services and CDC (2001) *PATCH, a Planned Approach to Community Health; Guide for the Local Coordinator*. Atlanta, GA.

- Van den Broucke, S. (2003) The role of academic research in health promotion capacity building. In *Interregional Consultancy on capacity building for Health Promotion WHO*, Bangkok.
- Van den Broucke, S., De Potter, B., Lenders, F. and Vinck, J. (1994) An assessment of methodological quality of health promotion project in Flanders *Archive Public Health*, 52, 261-277.
- Van den Broucke, S., Molleman, G. R. M., Broesskamp-Stone, U., Speller, V. and Saan, J. A. M. (2004) 'Getting evidence into practice': tools and processes for Health Promotion, special session of the GPHPE/European. In *Health 2004*. Melbourne.
- Veen, C., Vereijken, C., Driel, W. v. and Belien, M. (1994) *An instrument for analysing effectiveness studies on health promotion and health education*. International Union for Health Promotion and Education, Utrecht.
- Vermeulen, T. R. N., Kok, H. H., Van den Broucke, S., Molleman, G. R. M. and Bollars, C. (2005) Health Promotion Quality Assurance Self-Assessment tools; systematic review and qualitative content analysis *submitted*.
- Victorian Community Health Association (2004) *Quality Improvement Program Planning System (QIPPS)*. website www.qipps.com. Victorian Community Health Association, Melbourne
- VNG, GGD Nederland and NIGZ (2004) *Convenant versterking Gezondheidsbevordering*. VNG, Den Haag.
- Voorham, A. J. J. (2003) *Gezondheidsbevordering voor-en-door de doelgroep; theoretische onderbouwing en evaluatie bij migranten en ouderen*. GGD Rotterdam e.o., Rotterdam.
- Vries, H. de, Kuhlman, P. and Dijkstra, M. (1987) Persoonlijke effectiviteit: de derde variabele naast attitude en subjectieve norm als voorspeller van de gedragsintentie *Tijdschrift Gezondheidsbevordering*, 253-264.
- Vries, N.K. de (1993) Groepsbeslissingen. In Koele, P. and Pligt, J. v. d. (eds), *Bestissen en beoordelen*. Boom, Amsterdam, Meppel, p. 251-284.
- Vries, N. K. de (2000) *Het hart, de ruggegraat en de hersenpan; perspectieven op gezondheidsgedrag*. Universiteit van Maastricht, Maastricht.
- Wagenaar, W. A. (1987) Logisch voorwaarts en intelligent achterwaarts; modellen voor het stellen van diagnoses in de hulpverlening. In Ploeg, J. D. v. d. and Berg, P. M. v. d. (eds), *Besluitvorming en jeugdhulpverlening*. Acco, Leuven, Amersfoort, p. 9-18.
- Walburg, J. A. (1997) *Integrale kwaliteit in de gezondheidszorg*. Kluwer, Deventer.
- Woerkmom, C. v. and Meegeren, P. v. (Eds.) (1999) *Basisboek Communicatie en Verandering*. Boom, Amsterdam.

- Woodard, G. B. (2004) *Health Promotion Capacity checklists: a Workbook for Individual, Organisational, and Environmental Assessment*. Prairie region Health Promotion Research Centre, University of Saskatchewan, Saskatoon, Saskatchewan, Canada.
- Work Group on Health Promotion and Community Development (1995) *Community toolbox, website <http://ctb.ku.edu/>*. University of Kansas, Lawrence, Kansas.

Summary

This thesis discusses the systematic development of the second version of the Health Promotion Effectmanagement Instrument, called Preffi 2.0, and studies to assess its usefulness, reliability and validity. Preffi 2.0 is an effect management instrument intended to help health promotion specialists improve the effectiveness of their interventions by applying the Preffi criteria. These criteria reflect both research findings on effect predictors and practitioners' knowledge about effectiveness, including aspects like context and project management. The instrument can be applied to individual health promotion interventions, projects involving multiple interventions or entire programmes consisting of multiple projects. Preffi is primarily intended to be used by health promotion and prevention practitioners. New versions of the instrument will be periodically produced. The idea is to create a learning system that can shape the dynamic relation between researchers and practitioners.

There are still many opportunities to improve the effectiveness of health promotion interventions. To achieve such improvements, it is necessary not only to develop and disseminate model programmes, but also to stimulate local health promotion specialists to apply general principles and guidelines for effectiveness in their daily routine. The use of Preffi as an effect management instrument should help achieve this. The introductory chapter to this thesis discusses Preffi's status and role as an instrument within the broader concept of quality management, one element of which is effectiveness. Effectiveness is greatly influenced by the operational processes used in an intervention or programme.

The development of a second version of Preffi had to address a number of aspects. It was developed based on a model of the structure of planning processes. This model includes dimensions that contribute to effectiveness: programme development, the programme itself, implementation and evaluation. For each of these dimensions, we identified effect predictors based on considerations of content, project management and context. Together, these became the effect predictors, or criteria, that constitute Preffi 2.0.

The further development of the instrument will try to tie in as closely as possible with new insights about guideline development and the dissemination of innovations. The Preffi development programme has always been based on the users' perspective, which is why health promotion specialists are involved in all

developmental stages and decision moments. The second version of Preffi incorporates a more fully developed scientific rationale and devotes considerable attention to validity and reliability.

A recurrent topic in this thesis is the role and status of Preffi: is it a diagnostic instrument that helps and stimulates practitioners to improve their own projects, or is it intended as a screening instrument to assess the quality of projects for selection purposes? The instrument's developers emphasise Preffi's diagnostic purpose. This has required standards (norms) to be defined, against which users can assess their own work.

Since most of the chapters in this thesis concentrate on specific parts of the research and development programme that has led to Preffi 2.0, the introductory chapter surveys the entire Preffi programme, which started in 1994. As such, it discusses the developmental and implementation activities for Preffi 1.0, the development of Preffi 2.0 and the national and international activities that have been initiated after the launch of Preffi 2.0.

Chapters 2 and 3 discuss the experiences gained with the first version of Preffi over the period from 1995 to 2001. These experiences and the empirical findings of studies examining the 1997 – 1999 implementation programme formed the basis for Preffi 2.0.

Chapter 2 describes the experiences gained with Preffi 1.0. Preffi 1.0 was intended as an instrument to help practitioners apply available research and practical knowledge in their own everyday work routine. Its format was that of a checklist of key points that have to be addressed in developing and implementing health promotion interventions and projects. Practitioners had expressed a preference for a short, clearly structured list, which would serve not only to assess the quality of projects but especially to improve that quality.

The experience gained with Preffi 1.0 and studies examining its use showed that health promotion specialists are very much action-oriented and that they tend to make rapid and pragmatic decisions about the interventions to be used in their projects. They generally find it difficult to motivate their intervention choices on theoretical grounds, frequently lack insight into determinants and often fail to state their objectives in clear terms.

Using Preffi can help them look at their projects in a more systematic way, which implies a greater interest in theory and determinants, more precise descriptions of the objectives that a particular intervention should be able to

achieve for particular sections of the target group, and the setting of more realistic objectives in view of the contextual conditions. Preffi can help them review recent research findings, which is important because the average Dutch health promotion specialist completed his training more than eight years ago.

Studies of the implementation programme showed that users did not always find it easy to apply Preffi. They often wondered whether they had indeed met particular criteria, in other words, they wanted a standard against which they could assess their own work. They also said that using Preffi took a great deal of time, although this comment was made particularly by people who had only just started to work with the instrument.

Repeated studies assessing the use of Preffi in 1997, 1998 and 1999 showed that the health promotion specialists attached great value to an opportunity to analyse the effects of their projects, using a list that would be valid for all health promotion sectors. The practitioners regarded Preffi as an important, supportive and useful instrument. They tended to use it in a flexible manner, in the sense that they applied those parts of Preffi 1.0 that were relevant to their specific context. They used it particularly during the developmental stages of their projects. By the end of 1999, 96% of the Dutch health promotion specialists were aware of the existence of Preffi, and 35 – 40% of them were using it as their standard instrument.

The experiences gained with Preffi 1.0 indicated what improvements could be made to the instrument itself. In terms of its *content*, recent research findings had to be incorporated, and the instrument had to pay more attention to the theoretical underpinning of projects, the motivation of target groups, target group participation, efforts to make projects fit the culture of target groups, implementation strategies, the context and contextual conditions in which projects are implemented and the importance of the project manager as a decisive factor for success. All criteria had to be operationalised and provided with norms. In terms of the *format*, the new version had to reflect the cyclical and iterative nature of health promotion projects. As regards the *status and role* of the instrument, it not only had to target the practitioners themselves, but also had to stimulate support from researchers and institutional management. The new name Health Promotion Effect Management Instrument more clearly reflects Preffi's intention to identify conditions for effectiveness, rather than measuring effectiveness as such, as was sometimes erroneously thought.

Chapter 3 discusses the empirical research into the Preffi 1.0 implementation programme between 1997 and 1999. The Theory of Planned Behaviour was

used as the theoretical model to plan the implementation interventions and structure the research. The objectives of the Preffi 1.0 implementation programme were related to its dissemination, adoption, implementation and maintenance. The determinants of adoption and the transition to the implementation stage were attitude, social norm and self-efficacy, while the determinants of progress in the implementation process were the characteristics of the instrument, the social and political context, the 'sender' and 'receiver' of the innovation and the implementation strategy used.

Nationwide implementation interventions targeting all Dutch health promotion specialists aimed at raising awareness of the existence of Preffi, creating a favourable attitude and providing general assistance for its use. In addition to this general, nationwide strategy, there was also an intensified, five-day support programme, which was offered to a group of 64 health promotion specialists, who intended to use Preffi. The effects of the general, nationwide strategy were assessed in two independent samples of Dutch health promotion specialists (N=120 en 316, respectively) using a written questionnaire, while the effects of the intensified training programme were assessed in a cohort study. The nationwide activities raised the awareness of the existence of Preffi and resulted in a more favourable attitude towards the instrument. However, only those who had participated in the intensified training programme showed a significant increase in the use of Preffi.

The favourable attitude was influenced particularly by opinions about using a systematic approach and about the usefulness of the instrument. A striking finding was that institutional management had generally not stimulated health promotion specialists to use Preffi. In those cases where Preffi had been integrated into a team's quality assurance system (team embeddedness) this was found to be an important predictor of implementation, and in fact the only predictor of the transition from the action stage to the maintenance stage (OR=1.521).

The most important factor for the transition from the preparation stage to the action stage was the attitude scale relating to the 'usefulness of the instrument' (OR=1.989), while team embeddedness (OR=1.646) and self-efficacy (OR=1.328) also played important roles.

Chapter 4 discusses the systematic process used to develop Preffi 2.0, which was used to ensure the instrument's validity. Preffi 2.0 incorporates the recommendations for content, format and status that came out of the Preffi 1.0 programme. We tried to strengthen the scientific legitimacy of Preffi 2.0 by collab-

orating with the Prevention Research Centre at Radboud University Nijmegen and establishing a scientific advisory committee (WAR). In addition, we set up a practitioners' advisory committee (PAR) of Dutch health promotion specialists, that is, the instrument's primary users.

The development process started with an explicit plan for the way the draft version of Preffi 2.0 was produced, after which the content and validity of this version were examined by the WAR and PAR members. Finally, 35 experienced health promotion specialists were asked to assess the usefulness of a draft version. The chapter presents some of the results of this study and the conclusions we drew from them for the use of the instrument. The users taking part in the study generally gave favourable opinions, and the study yielded useful suggestions for adjustments to the instrument's content and lay-out, as well as its possible and preferred uses. These suggestions were incorporated in the definitive version of Preffi 2.0. In this version, Preffi 2.0 consists of a scoring form with 39 criteria, divided into 8 clusters. Each criterion is operationalised and provided with a norm which allows users to rate the degree to which a project meets that criterion as strong, moderate or weak. The assessor is stimulated to note specific suggestions for improving the project being assessed. The Preffi package includes a user manual and an elaborate explanatory guide that justifies the choice of criteria.

The chapter discusses the differences between Preffi 2.0 and the 1.0 version and gives some suggestions for using the instrument, such as the recommendation to assess projects together with colleagues and the recommendation not to assess a project purely on the basis of documents but to consult with the project manager. The study showed that users prefer to use Preffi for diagnostic purposes.

Chapter 5 reports on the study assessing the usefulness and reliability of a draft version of Preffi 2.0 among 35 health promotion specialists, each of whom was asked to assess two projects. The respondents also commented on the operationalisations and completed a questionnaire asking their opinion about the instrument. Supplementary interviews were held with 10 of them. The Preffi-based assessments of projects based on the various criteria were found to differentiate sufficiently between projects, as well as between the various criteria for a particular project. The assessors frequently made use of the 'not assessable' answering option, especially for the criteria newly introduced in Preffi 2.0 and for criteria from the clusters on contextual conditions and feasibility, implementation and evaluation. It was clear that it is difficult to assess projects merely on the basis of written project descriptions; it is useful to include in the

assessment procedure a consultation with the project manager, to obtain supplementary information.

We used generalisability theory to assess the reliability and accuracy of Preffi as an instrument. The study did not produce any conclusive evidence about reliability, as there was insufficient variance between the projects assessed. As regards accuracy, we concluded that sufficiently accurate assessments on the basis of criteria scores required about 4 assessors for the individual clusters and 2 for the project as a whole. A sufficiently accurate and reliable assessment of a project based on the overall marks for the various clusters and for the project as a whole would require between 25 and 40 assessors. The study provided valuable suggestions for further research that would allow the instrument's reliability to be more accurately established, as well as suggestions to increase the instrument's reliability, such as using fewer answering options, training users in the application of Preffi and clarifying the descriptions and instructions.

The respondents gave the draft version of Preffi 2.0 an overall rating of 7.7 out of 10. Most of them reported that Preffi was a valuable, complete, clear, well-structured and innovative instrument, but they also said it was not easy to apply. The respondents expected that applying the instrument to their own projects would enable them to be more critical of their own work, to systematically assess all aspects of their projects and to rapidly draw up a list of points that could be improved.

The team that developed Preffi 2.0 emphasises its role as a quality assurance instrument, intended to stimulate improvements. The instrument in its present form is not suitable as a screening tool, although it does provide a standard that specific health promotion projects should ideally be able to meet. The respondents agreed with this view and expressed a certain fear that third parties might use Preffi in an uncontrolled fashion as a screening instrument to assess projects.

Chapter 6 reports on a study to test the reliability of the definitive version of Preffi 2.0. It compared Preffi-based assessments of 20 projects by three practitioners with the intuitive assessments of the same projects they had given earlier and with assessments by three experts serving as an external criterion. The intuitive assessments involved assigning marks for eight general aspects of the projects, which corresponded to the clusters used in Preffi. The main hypothesis of the study was that the intuitive assessments by the practitioners would be less reliable and accurate than their Preffi-based assessments and than experts' assessments. This hypothesis was, on the whole, not confirmed. The assess-

ments by the experts turned out to be less reliable and accurate than the intuitive and Preffi-based assessments by the practitioners, and differed too much to be used as an external criterion. The practitioners' intuitive assessments and Preffi-based assessments did achieve acceptable reliability. Based on the criterion scores, sufficiently reliable and accurate assessments at the level of projects as a whole would require 2 assessors, while 3 would be needed at the level of individual clusters. This study also found that the overall marks given to individual clusters were less reliable and accurate.

The reliability and accuracy of the Preffi scores were better than those found in the earlier study of the draft version of Preffi 2.0. The changes introduced on the basis of the findings of that earlier study resulted in fewer assessors being required for accurate and reliable assessments of all criteria that had been adjusted.

The experts' assessments of the projects were on average the most critical. There was no difference in strictness between the practitioners' intuitive and Preffi-based assessments, although the respondents themselves had the impression that Preffi forced them to be more critical. The practitioners in this study also reported that they found it hard to assess many of the projects purely on the basis of written project descriptions and would prefer to have a consultation with the project manager included in the assessment procedure.

The study showed that different assessors had different perspectives and included different aspects in their assessments. The assessors themselves regarded this as inevitable and even as valuable. They thought that a consensus meeting between different assessors should be a standard element in the assessment procedure. The consensus meeting that was held as part of the study proved that consensus about individual projects was quickly achieved, based mostly on theoretical arguments about the quality of the projects.

Chapter 7 summarises and critically reviews the findings of the various studies reported on in this thesis. The overall conclusion is that Preffi 2.0 is a valid instrument using evidence-based principles and providing guidelines on effect management. The new version represents an improvement over the first version, while still remaining useful and feasible for practitioners. There is general agreement about the good content validity of the instrument. Our studies have shown how Preffi can best be used to reliably assess a project. Preffi allows the strong and weak points of projects and programmes to be adequately diagnosed. It also contributes to a synthesis of new research findings and practical experience, making them accessible to practitioners.

The Preffi team hopes to be able to produce a new version (Preffi 3.0) by 2007. To this end, the instrument's content and format would need to be further adjusted, incorporating new insights into ways of increasing its reliability and the steps required to further integrate a consistent use of Preffi in existing operational processes.

We would also recommend the inclusion of cost-effectiveness aspects in the new version. Preffi 3.0 should in any case have a digital, Internet-supported format. This might further increase the instrument's reliability, as it would allow more specific explanations and suggestions to be provided for each criterion, which may help users achieve more objective assessments. It will be necessary to continue to collect empirical data that could allow the use of Preffi to be further improved.

The consistent use of the instrument can be further stimulated by incorporating Preffi or its principles in the quality assurance system for health promotion which is being developed by the Foundation for the Harmonisation of External Quality Review in Health Care (HKZ) and is to start in 2006, as well as by integrating them into the operational processes and the various knowledge products provided by the NIGZ Centre for Knowledge and Quality Management.

The insights gained with Preffi are making important contributions to the current attempts to develop a European guideline for health promotion. Collaboration with colleagues from various European countries should improve the chances of success. All of these plans indicate that the further development of Preffi and efforts to improve the effectiveness of health promotion in the Netherlands require a permanent investment effort.

Samenvatting

Dit proefschrift behandelt de systematische ontwikkeling van de tweede versie van het Preventie Effectmanagement Instrument, de Preffi 2.0 en de onderzoeken naar de bruikbaarheid, betrouwbaarheid en validiteit van het instrument. De Preffi 2.0 is een instrument voor effectmanagement dat gezondheidsbevorderaars helpt om de effectiviteit van hun interventies te verbeteren door de Preffi-criteria en richtlijnen toe te passen op hun interventies. De criteria reflecteren zowel wetenschappelijke kennis over effectpredictoren als praktijkinzichten over effectiviteit, zoals aandacht voor context en projectmanagement. Het kan daarbij gaan om afzonderlijke interventies, projecten met meerdere interventies of programma's die meerdere projecten omvatten.

De primaire gebruiker van de Preffi is de praktisch werkende gezondheidsbevorderaar en preventiewerker. Periodiek zullen er nieuwe versies van de Preffi worden uitgegeven. Hiermee wordt beoogd een lerend systeem in het leven te roepen dat vorm kan geven aan de dynamische relatie tussen wetenschap en praktijk.

Er kan nog veel winst in effectiviteit geboekt worden bij interventies voor gezondheidsbevordering. Om dat te realiseren is het nodig om naast het ontwikkelen en verspreiden van modelprogramma's, te stimuleren dat (lokale) gezondheidsbevorderaars algemene principes en richtlijnen over effectiviteit in hun dagelijkse praktijk gaan toepassen. Voor dit laatste is de Preffi als instrument voor effectmanagement een bruikbaar hulpmiddel. In het inleidende hoofdstuk wordt de Preffi gepositioneerd als een instrument binnen het bredere concept van kwaliteitsmanagement, waarvan effectiviteit een onderdeel is. Effectiviteit wordt sterk bepaald door de werkprocessen die in een interventie of programma worden gevolgd.

Voor de ontwikkeling van een tweede versie van de Preffi is met een aantal aspecten rekening gehouden.

Er wordt uitgegaan van een model dat is gebaseerd op de structuur van het planingsproces. In het model zijn dimensies opgenomen die bijdragen aan effectiviteit: programma-ontwikkeling, het programma zelf, de implementatie ervan en evaluatie. Voor ieder van deze dimensies zijn effectpredictoren vastgesteld op basis van inhoudelijke en projectmatige/contextuele afwegingen. Te samen vormen ze de effectpredictoren of criteria waaruit de Preffi 2.0 is opgebouwd.

Bij de ontwikkeling en de opbouw van de Preffi wordt zoveel mogelijk aangesloten bij de inzichten uit richtlijnenontwikkeling en het verspreiden van innovaties.

Het gebruikersperspectief vormt vanaf het begin van het Preffi traject het vertrekpunt. Daarom worden gezondheidsbevorderaars betrokken bij alle ontwikkelingsstadia en keuzemomenten.

In de tweede versie van de Preffi is de wetenschappelijke onderbouwing van de Preffi verder uitgewerkt met veel aandacht voor validiteit en betrouwbaarheid. Een vaker terugkerend aandachtspunt in dit proefschrift is de functie die de Preffi kan spelen. Gaat het om een diagnostisch instrument dat de gebruiker helpt en stimuleert het eigen project te verbeteren of is de Preffi gericht op het beoordelen van projecten voor selectiedoeleinden? De ontwikkelaars willen de nadruk leggen op de diagnostische functie. Daarvoor is het noodzakelijk om een norm vast te stellen waartegen het eigen handelen kan worden afgezet.

Omdat de hoofdstukken in dit proefschrift zich toespitsen op een aantal onderdelen van het ontwikkelings- en onderzoekstraject voor de Preffi 2.0 wordt in het inleidend hoofdstuk het hele Preffi traject geschetst, zoals dat sinds 1994 is uitgevoerd. Ingegaan wordt op de ontwikkelings- en implementatieactiviteiten van de Preffi 1.0, de ontwikkeling van de Preffi 2.0 en de nationale en internationale activiteiten die na de lancering van de Preffi 2.0 zijn en worden uitgevoerd.

De hoofdstukken 2 en 3 gaan in op de ervaringen die zijn opgedaan met de eerste versie van de Preffi in de periode 1995-2001. Deze ervaringen en de empirische gegevens uit de studies naar het implementatietraject van 1997-1999 vormen de uitgangspunten voor de Preffi 2.0.

Hoofdstuk 2 beschrijft de ervaringen die met het Preventie Effectiviteits Instrument, de Preffi 1.0 zijn opgedaan.

De Preffi 1.0 wordt gepositioneerd als een instrument dat de praktijkwerker helpt om beschikbare wetenschappelijke en praktijk-gebaseerde kennis toe te passen in de eigen praktijk. Het heeft de vorm van een richtlijnenlijst met aandachtspunten waaraan gedacht moet worden bij de opzet en het verloop van interventies en projecten voor gezondheidsbevordering. Het was de wens van de gebruiker om de Preffi te beperken tot een korte overzichtelijke lijst, die niet alleen geschikt is voor het beoordelen maar vooral om tot verbeteringen te komen van een project.

De ervaringen en het onderzoek laten zien dat gezondheidsbevorderaars sterk actiegericht zijn en in hun projecten op pragmatische gronden snel tot een interventiekeuze komen. Gezondheidsbevorderaars vinden het moeilijk een interventiekeuze op theoretische gronden te motiveren. Het inzicht in determinanten ontbreekt nogal eens en doelen worden matig geformuleerd.

Het werken met de Preffi helpt om planmatiger naar een project te kijken. Dat uit zich in meer aandacht voor theorie en determinanten, nauwkeuriger

omschrijven welk deel van de doelen voor welk deel van de doelgroep met een interventie kan worden gerealiseerd en reële doelen gezien de randvoorwaarden. Het gebruik van Preffi helpt om recente wetenschappelijke inzichten te vergaren, omdat men gemiddeld al meer dan 8 jaar geleden de eigen opleiding heeft afgesloten.

Tijdens het implementatietraject bleek dat de Preffi niet altijd even makkelijk is toe te passen. Het is vooral de vraag wanneer je het nu goed hebt gedaan. Men mist een standaard waartegen men de eigen werkzaamheden kan afzetten. Daarnaast dacht men dat het toepassen van de Preffi veel tijd kost. Dat argument speelde vooral bij mensen die de Preffi nog nauwelijks gebruikt hebben. Uit de herhaalde metingen over het gebruik van de Preffi in 1997, 1998 en 1999 blijkt dat de gezondheidsbevorderaars het belangrijk vinden om de effecten van hun projecten te kunnen analyseren en dat met een lijst te doen die voor alle sectoren binnen gezondheidsbevordering geldt. De praktijkwerkers vinden de Preffi belangrijk ondersteunend en bruikbaar. Ze gebruiken het instrument flexibel. Die delen van de Preffi 1.0 worden toegepast die in een specifieke context relevant zijn en dan met name tijdens het opzetten van een project. Eind 1999 kende 96% van de gvo/preventiewerker het instrument en gebruikte 35-40% van hen de Preffi als standaardinstrument.

De ervaringen met de Preffi 1.0 maakten duidelijk hoe de Preffi zelf verbeterd kon worden.

Qua *inhoud* dienden nieuwe wetenschappelijke inzichten te worden verwerkt. Meer aandacht was gewenst voor de theoretische fundering van een project, de motivatie van de doelgroep, doelgroepparticipatie, afstemming op de cultuur, implementatiestrategieën, context, randvoorwaarden en de projectleider als een kritische succesfactor. Alle criteria zouden geoperationaliseerd en genormeerd moeten worden.

In de *vormgeving* zou aandacht moeten zijn voor het cyclisch en iteratieve karakter van gezondheidsbevorderingsprojecten.

De *positionering* diende zich niet alleen op de beroepsgroep zelf te mikken, maar ook op ondersteuning vanuit wetenschap en management. Met een naam als Preventie Effectmanagement Instrument wordt scherper de positionering neergezet dat de Preffi de voorwaarden voor effectiviteit in kaart wil brengen en niet de effectiviteit zelf meet, zoals men nog wel eens ten onrechte dacht.

Hoofdstuk 3 behandelt het empirische onderzoek naar het implementatietraject van de Preffi 1.0 dat tussen 1997 en 1999 heeft plaatsgevonden. The Theory of Planned Behaviour is gebruikt als theoretische model om de interventies te plannen en het onderzoek te structureren. In het implementatietraject worden doelen gesteld gericht op disseminatie, adoptie, implementatie en behoud van

het gebruik van de Preffi. Determinanten van de adoptie en de overgang naar de implementatiefase zijn de attitude, sociale norm en eigen effectiviteit. De determinanten van de voortgang in het implementatieproces zijn de karakteristieken van de Preffi, de sociaal-politieke context, de boodschapper en de ontvanger van de vernieuwing en de implementatiestrategie.

Landelijke implementatie-interventies voor alle gezondheidsbevorderaars waren gericht op het bekendmaken van de Preffi, creëren van een positieve houding en in algemene zin aanbieden van ondersteuning bij gebruik. Naast deze algemene nationale strategie is er ook een intensief 5-daags ondersteunings-programma aangeboden aan een gemotiveerde groep gezondheidsbevorderaars ($n=64$). De effecten van het landelijke aanbod zijn in twee onafhankelijke representatieve steekproeven van Nederlandse gezondheidsbevorderaars ($N=120$ en 316 , resp.) gemeten met een schriftelijke vragenlijst. De effecten van het intensievere trainingsprogramma zijn gemeten in een cohort studie. De landelijke activiteiten droegen bij tot een grotere bekendheid met en een (nog grotere) positieve houding ten opzichte van de Preffi. Maar een significante toename van het gebruik trad alleen op bij degenen die een intensief trainingsprogramma hadden gevolgd. De positieve attitude werd vooral beïnvloed door de opvattingen over een systematische benadering en de bruikbaarheid van het instrument. Opvallend was het ontbreken van een stimulans vanuit het management aan de gezondheidsbevorderaars om de Preffi te gebruiken. Daar waar de Preffi wel was ingebed in het kwaliteitssysteem van een team (teamverankering) bleek dit ook een belangrijke voorspeller van implementatie te zijn en zelfs de enige voorspeller van de overgang van de actie naar de behoud-fase ($OR=1,521$).

Voor de overgang tussen de preparatiefase naar de actie-fase bleek de attitude-schaal 'bruikbaarheid van het instrument' de belangrijkste factor te zijn ($OR=1,989$), terwijl ook teamverankering ($OR=1,646$) en eigen effectiviteit ($OR=1,328$) een belangrijke rol spelen.

Hoofdstuk 4 behandelt het systematische ontwikkelingsproces van de Preffi 2.0 dat de grondslag voor de validiteit van het instrument vormt. In de Preffi 2.0 worden de aanbevelingen uit het Preffi 1.0 traject over inhoud, normering, vorm en positionering verwerkt. Er is ingezet op een versterking van de wetenschappelijke legitimiteit van de Preffi 2.0 door een samenwerking aan te gaan met het Prevention Research Centre van de Radboud Universiteit Nijmegen en een Wetenschappelijke adviesraad (WAR) in te stellen. Er is ook een Praktijkadviesraad (PAR) opgezet met de primaire gebruikers, de gezondheidsbevorderaars in Nederland.

In het ontwikkelingsproces is eerst geëxpliciteerd hoe een conceptversie van de Preffi 2.0 tot stand is gekomen, vervolgens is de inhoud en de validiteit in verschillende overlegondes getoetst met de experts van de WAR en PAR. Tot slot

is over een concept-versie een bruikbaarheid-onderzoek gehouden onder 35 ervaren gezondheidsbevorderaars. Enkele resultaten van dat onderzoek en leerpunten voor het gebruik worden in dit hoofdstuk gepresenteerd. Het gebruikersonderzoek liet in het algemeen een positief oordeel zien en heeft suggesties opgeleverd over aanpassing van inhoud en lay-out en de mogelijke en wenselijke wijzen van gebruik. Deze zijn verwerkt in de definitieve versie van het Preventie Effectmanagement Instrument, de Preffi 2.0. De Preffi 2.0 bestaat uit een scoreformulier met 39 criteria verdeeld over acht clusters. Ieder criterium is geoperationaliseerd en van een normering voorzien, waarmee een score sterk, matig of zwak kan worden toegekend. De beoordelaar wordt gestimuleerd op het formulier een aantal concrete verbeterpunten te benoemen voor het beoordeelde project. Bij de Preffi hoort een gebruiksaanwijzing en een uitgebreide toelichting, waarin de keuze voor de criteria wordt verantwoord. De veranderingen in de Preffi 2.0 ten opzichte van de 1.0 versie worden besproken. Een aantal aanwijzingen voor het gebruik worden gegeven, zoals de aanbeveling om samen met anderen een project te beoordelen en niet alleen te beoordelen op basis van schriftelijke informatie, maar ervoor te zorgen ook altijd een gesprek met de projectleider te hebben. De gebruikers prefereren een gebruik van de Preffi voor diagnostische doeleinden.

Hoofdstuk 5 rapporteert het bruikbaarheidsonderzoek en het eerste betrouwbaarheidsonderzoek van de conceptversie van de Preffi 2.0 onder 35 gezondheidsbevorderaars, die ieder 2 projecten hebben beoordeeld met de concept-versie van de Preffi 2.0, de operationalisering en hebben becommentarieerd en een vragenlijst hebben ingevuld. Met 10 van hen is een aanvullend interview gehouden. De oordelen die met de Preffi worden gegeven over de verschillende criteria van projecten bleken voldoende te differentiëren tussen projecten en ook tussen criteria binnen projecten. De antwoordcategorie 'niet te beoordelen' werd vaak gebruikt, vooral bij de nieuwe Preffi 2.0 criteria en bij criteria uit de clusters randvoorwaarden en haalbaarheid, implementatie en evaluatie. Duidelijk is geworden dat een oordeel over een project moeilijk alleen op een projectbeschrijving kan worden gebaseerd. Het verdient aanbeveling ook een gesprek met de projectleider in de beoordelingsprocedure op te nemen om de ontbrekende informatie te krijgen.

Met de generaliseerbaarheidstheorie is de betrouwbaarheid en de nauwkeurigheid van de Preffi als instrument onderzocht. Over de betrouwbaarheid kan met deze studie geen afrondend oordeel worden gegeven, omdat er te weinig variantie tussen de projecten was. Er bleken op clusterniveau ongeveer 4 en voor het project als geheel 2 beoordelaars nodig te zijn om op basis van de criteriascores tot voldoende nauwkeurige oordelen te komen. Voor een nauwkeurig en betrouwbaar oordeel over een project op basis van de gegeven rapportcij-

fers over de clusters en het totale project zijn tussen de 25 en 40 beoordelaars nodig. Het onderzoek geeft waardevolle suggesties voor vervolgonderzoek om de betrouwbaarheid beter te kunnen vaststellen en suggesties om de betrouwbaarheid van het instrument te verhogen, zoals minder antwoordcategorieën, trainen in het gebruik van de Preffi, aanscherping van de omschrijvingen en instructies.

De conceptversie van de Preffi 2.0 krijgt van de respondenten gemiddeld een 7,7. De meerderheid beoordeelde de Preffi als waardevol, compleet, duidelijk, goed gestructureerd en innovatief, maar ook als niet eenvoudig. De respondenten verwachtten dat ze door toepassing van het instrument in hun projecten kritischer worden over hun eigen werk, eerder aan alle aspecten aandacht schenken en sneller zullen kunnen beschikken over een goed overzicht van verbeterpunten voor hun eigen projecten.

De ontwikkelaars leggen de nadruk op het gebruik van de Preffi als een kwaliteitsinstrument om verbeteringen te stimuleren. Het instrument is nog niet geschikt als beoordelingsinstrument, maar geeft wel een normatieve standaard, waarvan het wenselijk is dat deze gehaald wordt in concrete gezondheidsbevorderingsprojecten. De respondenten onderschrijven deze stellingname en zijn enigszins beducht voor een ongecontroleerd gebruik van de Preffi als beoordelingsinstrument van hun projecten door derden.

In hoofdstuk 6 wordt de betrouwbaarheid van de definitieve Preffi 2.0 versie getest. De oordelen over 20 projecten, zoals drie praktijkwerkers die met de Preffi hebben gegeven, worden vergeleken met intuïtieve oordelen die zij eerder hebben gegeven over dezelfde 20 projecten en met de oordelen van drie experts, die als een extern criterium dienen. De intuïtieve oordelen bestonden uit het toekennen van cijfers op 8 algemene aspecten van een project. Deze algemene aspecten vielen samen met de clusters uit de Preffi. De belangrijkste hypothese was dat de intuïtieve oordelen van de praktijkwerkers minder betrouwbaar en nauwkeurig zouden zijn dan de op de Preffi gebaseerde oordelen en de oordelen van de experts. In het algemeen konden we deze hypothese niet bevestigen. De oordelen van de drie experts waren minder betrouwbaar en precies dan de intuïtieve en de Preffi-oordelen van de praktijkwerkers en ze verschilden te zeer van elkaar om als extern criterium te kunnen dienen. De intuïtieve oordelen en de Preffi oordelen van de praktijkwerkers behaalden acceptabele niveaus van betrouwbaarheid. Gebaseerd op de criteria-scores zijn op het niveau van het hele project 2 beoordelaars nodig om een voldoende betrouwbaar en precies oordeel te krijgen, terwijl op clusterniveau er 3 beoordelaars nodig zijn. Ook in dit onderzoek bleek dat de rapportcijfers die per cluster worden gegeven minder betrouwbaar zijn en een grotere onnauwkeurigheid kennen.

De betrouwbaarheid en de nauwkeurigheid van de Preffi-scores zijn beter dan

die uit het onderzoek naar de conceptversie van de Preffi 2.0. De veranderingen die zijn doorgevoerd naar aanleiding van dat onderzoek hebben er toe geleid dat voor alle aangepaste criteria minder beoordelaars nodig zijn om tot een voldoende nauwkeurig en betrouwbaar oordeel te komen.

De experts oordelen gemiddeld het meest streng over de projecten. Voor de praktijkwerkers maakte het niet uit of men met of zonder de Preffi een oordeel uitsprak, al had men wel het idee met de Preffi strenger te oordelen. In veel gevallen vond men ook in dit onderzoek het lastig alleen op basis een schriftelijke rapportage tot een oordeel te komen en had men graag ook een gesprek met de projectleider in de procedure opgenomen gezien.

Het onderzoek maakte duidelijk dat beoordelaars verschillende perspectieven gebruiken en verschillende aspecten in hun beoordeling betrekken. Dit werd door de beoordelaars als onvermijdelijk en zelfs nuttig beschouwd. Ze vonden dat een discussie tussen de beoordelaars een vast onderdeel van de procedure moet worden om tot consensus te komen. Tijdens een consensus bijeenkomst bleek dat men snel tot consensus over de projecten kon komen, waarbij vooral inhoudelijke argumenten over de kwaliteit van de projecten de doorslag gaven.

In hoofdstuk 7 worden de resultaten van de verschillende onderzoeken waarover is gerapporteerd in dit proefschrift samengevat en kritisch besproken. De afrondende conclusie is dat Preffi 2.0 een valide instrument is met evidence based principes en richtlijnen voor effectmanagement. De tweede versie is verder en beter ontwikkeld dan de eerste versie en blijkt voor de praktijk bruikbaar en behapbaar te blijven. De inhoudsvaliditeit van de Preffi wordt breed onderschreven. Het onderzoek heeft duidelijk gemaakt op welke wijze op een verantwoorde manier een betrouwbaar oordeel over een project gegeven kan worden met de Preffi. De Preffi geeft een goede diagnose waar projecten en programma's goed en minder goed in zijn. De Preffi levert een bijdrage aan het synthetiseren van nieuwe wetenschappelijke en praktische kennis en het toegankelijk maken daarvan voor de praktijk.

Er wordt voorgesteld om in 2007 met een nieuwe Preffi 3.0 te komen. Inhoud en vorm van de Preffi moeten dan weer verder uitgewerkt worden. Nieuwe inzichten om de betrouwbaarheid te verhogen kunnen daarin worden opgenomen en er kan rekening gehouden worden met verdere stappen die nodig zijn om het structurele gebruik van de Preffi in te bedden in bestaande werkprocessen.

Er wordt gepleit om in een volgende versie de kosten-effectiviteit op te nemen. De vorm zal in ieder geval digitaal zijn en via internet ondersteund moeten gaan worden. Dat kan leiden tot een verhoging van de betrouwbaarheid omdat op die manier een specifiekere uitleg en suggesties per criterium gegeven kunnen worden die helpend kunnen zijn om een oordeel te objectiveren. Empirisch

materiaal over het beoordelingen van projecten dient men te blijven verzamelen om daarvan te leren en zo het gebruik van de Preffi verder aan te scherpen. Het structurele gebruik van de Preffi kan verder vergroot worden door de Preffi of de principes daaruit in te bedden in het HKZ-kwaliteitssysteem voor gezondheidsbevordering dat in 2006 van start zal gaan en het in te bouwen in de eigen werkpraktijk en de verschillende kennisproducten van het NIGZ Centrum voor Kennis en Kwaliteit.

De inzichten uit de Preffi leveren een belangrijke bijdrage aan een te ontwikkelen Europese richtlijn voor gezondheidsbevordering. Samenwerking met collega's uit andere landen is een kansrijke ontwikkeling. Al deze plannen maken duidelijk dat de ontwikkeling van de Preffi en de bevordering van de effectiviteit van de gezondheidsbevordering in Nederland een blijvende investering vragen.

Curriculum vitae

Gerardus Richard Maria Molleman was born in Lichtenvoorde on 5 September 1954. He received his secondary school diploma ("*Gymnasium β*") at Thorbecke Scholengemeenschap in Arnhem in 1974. He then studied psychology at Nijmegen University, specialising in clinical psychology and particularly in mental health promotion.

From 1983 to 1994, he was head of the Department of Prevention and Education at the *Gelders Centrum voor Verslavingszorg* (Gelderland centre for addiction care) in Arnhem.

He has always taken a special interest in issues of quality management, professional development, and the continued development of the discipline of health promotion. In 1994, he started his own consultancy agency, called Bureau Maat, engaging in training and research activities and offering consultations and advice on strategic policy development and interim management for a variety of health promotion organisations.

Since 1994, he has been working for the Netherlands Institute for Health Promotion and Disease Prevention (NIGZ), where he has had a full-time position since 1999. His first task at NIGZ was to help develop a set of guidelines for practitioners, which later led to the development of Preffi. From 1996 to 1999 he led the professional development team, after which he was the director of the department of local health promotion. Since 2003, he has been the director of the NIGZ Centre for Knowledge and Quality Management. Throughout his NIGZ career, he remained responsible for the further development and implementation of Preffi.

Gerard Molleman has served on the boards of many associations. He was chairman of the Nederlandse Vereniging voor Preventie en GVO (Dutch association for prevention and health promotion, NVPG) from 1990 to 1993 and chairman of NVPG's *Wetenschappelijke Stichting* (research association) from 1995 to 2002. Since 2001, he has been the chairman of NVPG's professional development board.

Gerard Molleman represents NIGZ in EuroHealthNet and represents the Netherlands in the Working Party for Lifestyle (& other health determinants) of the European Commission's Health Information programme in Luxembourg. He is contract holder for the European project 'Getting evidence into practice', which is funded by the European Commission.

Appendix 1 Preffi 1.0, scoring form

Health PROMotion Effectiveness Fostering Instrument (PREFFI 1.E)	Assessment		Improvement scenarios	
	How done	Conclusion	How done	Priorities
Conditions				
0. Context <i>Organisational/political</i>				
Analysis				
1. Nature and scope of problem <i>a. scale of the problem</i> <i>b. seriousness of the problem</i> <i>c. material and immaterial cost resulting from problem</i> <i>d. problem-perception of relevant actors</i> 2. (Behavioural) determinants <i>a. contribution of determinant(s) to problem</i> <i>b. changeability of determinant(s)</i>				
Choices <i>(Targetgroup x Target x Intervention-Combination = TTIC's)</i>				
3. Targetgroup <i>a. general characteristics of targetgroup</i> <i>b. specific characteristics of segments of targetgroup</i> <i>c. inventarisation of needs, wishes and options of (segments of) targetgroup</i> 4. Target <i>a. targets are measurable</i> <i>b. targets are realistic/feasible</i> <i>c. targets are in balance with capacities/context</i> <i>d. targets specify aspects of problem tackled and those beyond scope</i> 5. Interventions <i>a. realistic intervention in balance with capacities/context</i> <i>b. choice of intervention accounted for in relation to target/ targetgroup</i> <i>c. intervention already proven to be effective</i> <i>d. intervention as efficient as possible</i> <i>e. intervention part of comprehensive strategy</i> <i>f. targetgroup involved in intervention choice and execution</i>				

Criteria	Assessment		Improvement scenarios	
	How done	Conclusion	How done	Priorities
<ul style="list-style-type: none"> <i>g. attention for side-effects</i> <i>h. anticipate on potential bottle-necks</i> <i>i. post-intervention follow-up (relapse and continuity)</i> <p>6. Effective elements build into intervention</p> <ul style="list-style-type: none"> <i>a. personalised approach</i> <i>b. feedback on effect</i> <i>c. positive behaviour rewarded</i> <i>d. remove barriers for preferred behaviour</i> <i>e. orchestrate social support</i> <i>f. train personal and social skills</i> <i>g. build in follow-up</i> <i>h. use of intervention-mix</i> 				
Management of the project				
<p>7. Comprehensive plan includes:</p> <ul style="list-style-type: none"> <i>a. responsibilities and roles</i> <i>b. human and financial resources</i> <i>c. timescale, milestones, actions</i> <i>d. documents and go/no go decisions</i> <i>e. involvement and contracts actors</i> <i>f. lines of communication</i> <i>g. requirements for qualified and adequate personnel</i> 				
Implementation of the intervention				
<p>8. Pre-test</p> <ul style="list-style-type: none"> <i>a. intervention[methods] tested and revised</i> <p>9. Execution</p> <ul style="list-style-type: none"> <i>a. organise publicity on the intervention</i> <i>b. options to adapt intervention to local circumstances</i> <i>c. availability of user-friendly scenarios</i> <i>d. make intermediaries aware of ad- and disadvantages</i> <i>e. training of intermediaries</i> <i>f. availability of intermediaries support</i> <i>g. organisation of follow-up events</i> <i>h. guaranteed continuity of intervention</i> <p>10. Evaluation</p> <ul style="list-style-type: none"> <i>a. process-evaluation</i> <i>b. effect-evaluation</i> 				

Appendix 2 Calculating G and SEM in generalisability theory

Measurements usually involve one discrimination facet (e.g. projects or persons). If one instrumentation facet is involved (and therefore also one discrimination facet), Cronbach's Alpha is a suitable measure of reliability. But Cronbach's Alpha cannot be used in a situation with several instrumentation facets or discrimination facets; such cases require the use of generalisability theory. It could be said that Cronbach's Alpha is a special case of generalisability theory.

In the usual cases, that is, without nesting or fixed facets, the calculations are as follows. The variance components of the facets and their interactions are used to estimate the true score variance ($Var(T)$) and the error variance ($Var(E)$). The true score variance consists of the variance in the discrimination facets and their interactions with each other. The error score variance consists of the variance in the instrumentation facets, their interactions with each other and their interactions with the discrimination facets. The G coefficient or reliability coefficient is then calculated as follows:

$$G = \frac{var(T)}{Var(T) + Var(E)}$$

The Standard Error of Measurement is the square root of the error variance:

$$SEM = \sqrt{Var(E)}$$

If there are fixed facets, certain error score facets are not included. If there are nested facets, certain main effects and interactions cannot be analytically distinguished. The method is, however, still based on combining variance components into estimates of $Var(T)$ and $Var(E)$.

The present study involved the following variance components: projects (p), assessors (res) and clusters/items (i) and their interactions. The discrimination facet relates to the projects. The items/clusters are regarded as exhaustive and hence as fixed facets. $Var(T)$ thus consists of the variance component of the projects and the projects*items interaction. All other components are included in $Var(E)$.

Since the number of assessors affects the magnitude of the error, manipulating this number allows us to calculate how many assessors are needed to achieve a sufficiently reliable assessment ($SEM < .26$). This is analogous to the Spearman-Brown prophecy formula for homogeneous test lengthening or shortening in classical test theory.